

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:41:06 ; Search time 42.48 Seconds  
(without alignments)  
476.036 Million cell updates/sec

Title: US-09-911-667A-4  
Perfect score: 1429  
Sequence: 1 EDVAEAIQFGKFWKLSVLT.....WFSNFSYGLVLLTTELPQ 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1429	100.0	548	AAV44633	Human organic cation transporter-like protein (OCT1p)
2	342.5	24.0	439	AA876766	Corynebacterium gl
3	342.5	24.0	448	AA876766	C glutamicum prote
4	338	23.7	742	AA94977	Human secreted pro
5	283.5	19.8	542	AA847274	HOAT3. Homo sapi
6	279.5	19.6	542	AA92902	Human cerebral org
7	278	19.5	536	AA92903	Rat cerebral organ
8	278	19.5	556	AA876766	Rat OCT1 protein.
9	271.5	19.0	535	AA51249	Rat liver anion tr
10	270.5	18.9	537	AAW4195	Mouse osteoclast t
11	268.5	18.8	561	AAW44196	Human osteoclast t

12	256.5	17.9	540	22	AA849401	Murine organic ani
13	255	17.8	554	19	AA84538	Human liver cell c
14	246.5	17.2	399	22	AA876717	Corynebacterium gl
15	246.5	17.2	431	22	AA876717	C glutamicum prote
16	246.5	17.2	607	21	AA812131	Hydrophobic domain
17	245	17.1	550	21	AA844278	Human organic anio
18	245	17.1	550	22	AA844278	HOAT1. Homo sapi
19	245	17.1	550	22	AA844278	Human organic anio
20	243	17.0	530	21	AA829626	Cat flea HMT synap
21	243	17.0	550	22	AA829626	Human protein havi
22	243	17.0	550	22	AA829626	Human organic anio
23	240.5	16.8	538	22	AA847273	HOAT2B. Homo sapi
24	240.5	16.8	546	22	AA847273	HOAT2A. Homo sapi
25	237.5	16.6	548	21	AA808823	A human organic an
26	236.5	16.6	551	20	AA808823	A human organic an
27	231	16.2	551	21	AA808824	A protein with cat
28	231	16.2	557	20	AA808824	Human carnitine tr
29	231	16.2	557	21	AA83929	Mouse organic anio
30	228.5	16.0	545	22	AA836553	A protein with cat
31	219	15.3	557	20	AA801652	Mouse OCTN2 amino
32	219	15.3	557	21	AA820580	Mouse carnitine tr
33	219	15.3	557	21	AA83930	Mouse OCTN3 protei
34	217.5	15.2	564	21	AA820578	C glutamicum prote
35	213.5	14.9	452	22	AA83094	Corynebacterium gl
36	208.5	14.6	446	22	AA876810	C glutamicum prote
37	208.5	14.6	475	22	AA876810	Arabidopsis thalia
38	200	14.0	522	21	AA816798	Arabidopsis thalia
39	200	14.0	522	21	AA852315	Arabidopsis thalia
40	199.5	14.0	541	22	AA847276	HOAT5. Homo sapi
41	196	13.7	491	21	AA816799	Arabidopsis thalia
42	196	13.7	491	21	AA852316	Arabidopsis thalia
43	195	13.6	578	22	AA806571	Human protein havi
44	193	13.5	551	20	AA801649	A protein with cat
45	191.5	13.4	473	21	AA816800	Arabidopsis thalia

#### ALIGNMENTS

RESULT 1  
AAV44633  
ID AAV44633 standard; Protein; 548 AA.  
AC AAV44633;  
XX  
DT 07-APR-2000 (first entry)  
XX Human organic cation transporter-like protein (OCT1p).  
DE Human; organic cation transporter-like protein; OCT1p; transporter;  
transmembrane; norepinephrine; neuroprotective; neuroleptic; anticonvulsant;  
antiParkinsonian; antidepressant; cellular process; cell proliferation;  
screen; treatment; prevention; diagnosis; neurodegenerative disorder;  
Alzheimer; Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;  
CNS disorder; central nervous system; schizophrenia; depression;  
behavioural; sleep disorder; Alzheimer's; eating disorder.  
XX Homo sapiens.  
XX  
XX  
XX Key Location/Qualifiers  
FT Domain 1..85  
FT Domain /label= Cytoplasmic\_domain  
FT Domain 141..154  
FT Domain /label= Cytoplasmic\_domain  
FT Domain 199..208  
FT Domain /label= Cytoplasmic\_domain  
FT Domain 259..314  
FT Domain /label= Cytoplasmic\_domain  
FT Domain 395..402  
FT Domain /label= Cytoplasmic\_domain  
FT Domain 448..457  
FT Domain /label= Cytoplasmic\_domain  
FT Domain 511..548





Db 34 ssglwaldmdvglisfymaalathwgsptetsilsgifvgmaigsaglladklgr 93  
 QY 80 KTGKLSVLTWLYGCI--LSAFAPVYSWILVLRGLVGGGG-VPOSVTLVYAEFLPWKA 135  
 Db 94 r--qvfalsllvvgatgasalsvslamlnalrfvvgiglaeipvastlissetprkv 150  
 QY 136 RAKCILLIEVMAIGTVEVVLAVFVMPSL--GWRWLLILSAVPLLLFAVLFCFWLPESAR 193  
 Db 151 rgrmvvileatwalgwlnaaivgtffvagsdngwrwalagcvpaivayvrglipesvr 210  
 QY 194 YDVLSSNOEKA---LATLKRITATENGAMPPLGKLIISROEDRGKMRDLFTPHFRWTTLLL 250  
 Db 211 flekgrhdeaealvvsfeaaaaegkaadattavvhdnaegsv-siwsaalrkrvtal 269  
 QY 251 WPIFNSAFSYG 263  
 Db 270 wivwfcinslyyg 282

## RESULT 4

AA94977  
 ID AAY94977 standard; Protein: 742 AA.

XX AC AAY94977;

XX DT 16-JUN-2000 (first entry)

XX DE Human secreted protein clone as180\_1 protein sequence SEQ ID NO:160.

XX KW Human; secreted protein; immunestimulant; immunosuppressant; virucide;  
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;  
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;  
 KW antihypertoid; immune deficiency; severe combined immunodeficiency; SCID;  
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;  
 KW connective tissue disease; multiple sclerosis; erythematosis;  
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;  
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;  
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;  
 KW autoimmune inflammatory eye disease; allergy.

XX OS Homo sapiens.

XX PN WO200009552-A1.

XX PD 24-FEB-2000.

XX PF 13-AUG-1999; 99WO-US18298.

XX PR 14-AUG-1998; 98US-0096622.

XX PR 17-AUG-1998; 98US-0096815.

XX PR 04-SEP-1998; 98US-0099229.

XX PR 23-OCT-1998; 98US-0105368.

XX PR 08-JAN-1999; 99US-0115234.

XX PR 12-FEB-1999; 99US-0119931.

XX PR 18-FEB-1999; 99US-0120575.

XX PR 30-APR-1999; 99US-0132020.

XX PR 11-AUG-1999; 99US-0096622.

XX PA (GEM) GENETICS INST INC.

XX PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

XX PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

XX PI Wong GG, Clark HF, Fechtel K;

XX DR WPI; 2000-205979/18.

XX PT New polynucleotides encoding secreted proteins, which may have e.g.

XX PT nutritional, chemokine, immune stimulating or suppressing.

XX PT hematopoiesis regulating, tissue growth, activin/inhibin

XX PS antiinflammatory or tumor inhibition activity

XX CC AAA16618 to AAA16697 encode the human secreted proteins given in  
 CC AAY94980, isolated from human adult brain, adult thyroid,  
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,  
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,  
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,  
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are  
 CC predicted to have biological activities which would make them suitable  
 CC for treating, preventing or ameliorating medical conditions in humans  
 CC and animals. The polynucleotides can be used as markers for tissues in  
 CC which the protein is preferentially expressed, as molecular weight  
 CC markers on Southern gels, and as chromosome markers or tags to identify  
 CC chromosomes or to map gene positions. The proteins can be used in the  
 CC treatment of immune deficiencies and disorders, such as severe combined  
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other  
 CC infections. These infections include human immunodeficiency virus (HIV),  
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and  
 CC candidiasis. The proteins can be used to treat autoimmune disorders such  
 CC as connective tissue disease, multiple sclerosis, systemic lupus  
 CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,  
 CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent  
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and  
 CC autoimmune inflammatory eye disease. The proteins can also be used to  
 CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent  
 CC probes for the human secreted proteins from the present invention.

XX SQ Sequence 742 AA;

Query Match 23.7%; Score 338; DB 21; Length 742;

Best Local Similarity 27.4%; Pred. No. 1.7e-29;

Matches 86; Conservative 54; Mismatches 126; Indels 48; Gaps 5;

QY 1 EDAYEALGFCFKFQWKLVTGLAWMADAMWILSILAPQLHCEWRLPWSQVALLTSVVF 60

Db 153 ealirecghrfqwtlyfvlglalmadgvefvvvgfvlpsaekmcldsdnkgmlgliv 212

QY 61 VGMSSSTLNGNISDQYGRKTGLKISVLWLYYGLISAFAPVYSWILVLRGLVGGIGG- 119

Db 213 lgmvmvgaflwggldrlgrtqcllisvsvsfafsfvgyggtfllcrllsgvgiggs 272

QY 120 VPOSVTLYAEFLPMKARAKCILLIEVFWALCTVFEVVLAVFVMPSLG----- 166

Db 273 ipivfsyfeqlaqekrehlswlcmfmvggyaaamawailphygwsfqmgsayqfhs 332

QY 167 WRMLLILSAVPLLLFAVLFCFWLPESARYDVLISGNOEKAIATLKRIATEN---CAP--M 219

Db 333 wrvflvcafpvfaigtlttqpesprfllengkhdeamvmlkvhdtnmrakghpervf 392

QY 220 PLGKLIISROEDR-----GKMRDLFTPHFRWTTLLLWF 252

Db 393 svthikthqedelieqlsdgtgwygrgvralslsggvwgnflscfgpyrritlmmmg 452

QY 253 IFFSNAFSYGVLVL 266

Db 453 wvftmsfsyygltcv 466

RESULT 5

AAB47274

ID AAB47274 standard; Protein: 542 AA.

XX AC AAB47274;

XX DT 06-AUG-2001 (first entry)

XX DE hOAT3.

XX KW Human; organic anion transporter; hOAT; liver; kidney;

XX KW membrane protein; transport; organic anion; splice variant.

XX OS Homo sapiens.

XX PS Claim 169; Page 614-616; 641pp; English.





RESULT 7  
 AAY92903  
 ID AAY92903 standard; Protein: 536 AA.  
 XX  
 AC AAY92903;  
 XX  
 DT 26-SEP-2000 (first entry)  
 XX  
 DE Rat cerebral organic anion transporter OAT3 protein.  
 XX  
 DE Rat; organic anion transporter protein; OAT3; cerebral tissue.  
 KW  
 XX  
 OS Rattus sp.  
 PN WO200017237-A1.  
 XX  
 PD 30-MAR-2000.  
 XX  
 PF 20-SEP-1999; 99WO-JP05120.  
 XX  
 PR 18-SEP-1998; 98JP-0265126.  
 XX  
 PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 PI Endou H, Sekine T, Kusuvara H;  
 XX  
 DR WPI; 2000-283546/24.  
 DR N-PSDB; AAA11147.  
 XX  
 PT Organic anion transporter protein OAT3 expressed in cerebral tissue,  
 PT antibodies to it and gene encoding it for study of cerebral transport  
 PT of anions including drug molecules -  
 XX  
 PS Claim 2; Page45-46; 48pp; Japanese.  
 XX  
 CC This sequence represents a rat organic anion transporter protein OAT3.  
 CC The sequence is used to regulate the transport of anions in cerebral  
 CC tissue, including drug molecules, and regulation of this transport.  
 XX  
 SQ Sequence 536 AA;  
 XX  
 Query Match 19.5%; Score 278; DB 21; Length 536;  
 Best Local Similarity 33.68; Pred. No. 7.2e-23;  
 Matches 75; Conservative 37; Mismatches 95; Indels 16; Gaps 6;  
 QY 57 SVVFGMMSSSTLGNISDQYGRKTGLKISVLWTLYGILSAFAPVYSWILVNLGLVGF 116  
 DQ 127 SIFMAGILVGPVIGELSDRGRKILTSWYLMIAASGSAFSPVYMIFRILGCS 186  
 QY 117 IGVDPQS-VTLVYAEPLPKAKACILLIEVFVWIAICTVFEVVLAVFVMPSLGMRWLLILSA 175  
 DB 187 IGSITSLTVINVEVPTSMRAISSTSGYCYTIGFILGLA-YAIPQ--WRWLQLTSS 243  
 QY 176 VPLLFAVLCEWLPESARYDVLGSGNOEKAIATLKRAT-----ENCAPMLGKLLISRQE 230  
 DB 244 APFFILSLWVPSIRVWISGYSKAIKTLQVATFNKGKKEGKTLIEELKFNQK 303  
 QY 231 D-----RCKMRDLF-TPHFRWTTLLWFIFSNFASYGVL 266  
 DB 304 DITSKVKYGISDLFRVSILRVTFCLSLAWFSTGFAYYSIAM 346  
 RESULT 8  
 AAR77676  
 ID AAR77676 standard; Protein: 556 AA.  
 XX  
 AC AAR77676;  
 XX  
 DT 10-SEP-1996 (first entry)  
 XX  
 DE Rat OCT-1 protein.  
 XX

KW Rat; OCT-1; transporter protein; cationic; xenobiotic; pharmaceutical;  
 blood; liver; kidney; epithelial cell; intestine; tetraethylammonium;  
 KW proximal renal tubule cell; intestine; enterocyte; transgenic; renal;  
 XX biliary; excretion; resorption; modulator; uptake.  
 OS Rattus rattus.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 20..46  
 FT /note= "transmembrane domain"  
 FT Modified-site 71  
 FT /note= "putative N-glycosylation site"  
 FT Modified-site 97  
 FT /note= "putative N-glycosylation site"  
 FT Modified-site 113  
 FT /note= "putative N-glycosylation site"  
 FT Domain 154..171  
 FT /note= "transmembrane domain"  
 FT Domain 178..197  
 FT /note= "transmembrane domain"  
 FT Domain 243..260  
 FT /note= "transmembrane domain"  
 FT Domain 267..283  
 FT /note= "transmembrane domain"  
 FT Domain 350..366  
 FT /note= "transmembrane domain"  
 FT Domain 380..398  
 FT /note= "transmembrane domain"  
 FT Domain 406..425  
 FT /note= "transmembrane domain"  
 FT Modified-site 432  
 FT /note= "putative N-glycosylation site"  
 FT Domain 435..452  
 FT /note= "transmembrane domain"  
 FT Domain 469..485  
 FT /note= "transmembrane domain"  
 FT Domain 494..514  
 FT /note= "transmembrane domain"  
 XX  
 PN DE4424577-A1.  
 XX  
 PD 18-JAN-1996.  
 XX  
 PF 13-JUL-1994; 94DE-4424577.  
 XX  
 PR 13-JUL-1994; 94DE-4424577.  
 XX  
 PA (FARH ) HOECHST AG.  
 XX  
 PI Gorboulev V, Gruendeman D, Koepsell H;  
 XX WPI; 1996-069422/08.  
 DR N-PSDB; AAT08702.  
 XX  
 PT Transporter protein for cationic xenobiotic(s) and pharmaceuticals,  
 PT and related DNA and transformed cells - used e.g. to assess  
 PT excretion and resorption of cationic cpds.  
 XX  
 PS Claim 1; Fig 2A1; 13pp; German.  
 XX  
 CC This is the amino acid sequence of the rat OCT-1 protein, a new  
 CC transporter protein able to transport cationic xenobiotics and  
 CC pharmaceuticals from the blood into liver or kidney epithelial  
 CC cells or from the intestine. The gene was isolated by injecting  
 CC a rat kidney gene library into Xenopus laevis oocyte and isolating  
 CC clones conferring uptake of 14C-tetraethylammonium. One clone  
 CC designated OCT-1 was isolated. Expression of the gene was detect  
 CC in proximal renal tubule cells, in liver epithelial cell and in  
 CC intestinal enterocytes. The DNA can be used to generate transgenic  
 CC cells for use in vitro test for renal/biliary excretion or  
 CC intestinal resorption of xenobiotics and pharmaceuticals. The protein  
 CC or cells expressing it can also be used to isolate modulators that  
 CC block uptake of pharmaceutical by the renal tubules.



Db 65 gpngkpecklrfvhlpnaslpndtqgatepcldgwiynstrdtivtewdlvcgsnklkem 124  
 QY 55 LTVSVFVGMSSSTLWGNISDOYGRKTGLKISVLWTLTYGILSAFAPVYSWILVRLGLVG 114  
 Db 125 aqsfmagilvgpvfgeisdrgfgrkpltwsllylaasgssafpsltvymifrlcig 184  
 QY 115 FGIGGVPOS-VTLYAEFLPMKARAKCILLIEVFMAICTVFEVVLAVFVMPISLG-----WR 168  
 Db 185 csisglstlilnvewpstraissttligycytigq-----filpglayavpqr 236  
 QY 169 WLLILSAVPLLLFAVLCFWLPESARYDVLGSGNOEKATLKRAT-----ENGAPMPLGK 223  
 Db 237 wlqisvsaaffilswvpesirwlvjsgkfsralktlqrvtatfngkkeegekitvee 296  
 QY 224 LIISROED-----RGKMRDLF-TPHFRWTTLLWFWFNSAFSYYGLVL 266  
 Db 297 lknldqkdsakvkylgslfrvsilrrvtfclsawfatgfayyslam 346

## RESULT 11

AAW44196  
 ID AAW44196 standard; Protein; 561 AA.  
 XX  
 AC AAW44196;  
 XX  
 DT 14-MAY-1998 (first entry)  
 XX  
 DE Human osteoclast transporter protein.  
 XX  
 KW Human; osteoclast transporter protein; osteoporosis; osteopetrosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9742321-A1.  
 XX  
 PD 13-NOV-1997.  
 XX  
 PF 09-MAY-1997; 97WO-US07856.  
 XX  
 PR 09-MAY-1996; 96US-0647397.  
 XX  
 PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.  
 XX  
 PI Beier DR, Brady KP;  
 XX  
 DR WPI; 1997-558983/51.  
 DR N-PSDB; AAV12393.  
 XX  
 PT Novel human and mouse osteoclast transporter proteins - useful for  
 PT isolating agents that decrease osteoclast activity, for treatment of  
 PT osteoporosis and osteopetrosis  
 XX  
 PS Disclosure: Page 35-39; 52pp; English.  
 XX  
 CC The present sequence represents human osteoclast transporter protein.  
 CC Agents that bind the nucleic acid encoding the osteoclast transporter  
 CC protein can be used to decrease its expression, thereby decreasing  
 CC osteoclast activity. The agents can be used to treat osteoporosis  
 CC and osteopetrosis.  
 XX  
 SQ Sequence 561 AA;

Query Match 18.8%; Score 268.5; DB 18; Length 561;  
 Best Local Similarity 24.6%; Pred. No. 9.2e-22;  
 Matches 86; Conservative 49; Mismatches 122; Indels 93; Gaps 12;  
 QY 2 DAVEAIG-FGRFOWKLSVLTGLAW--MADAMEMMLSLAPQLHCE-----WRLP- 48  
 Db 5 eildrvsgmhgfihvailgilnmanhnlqiftaatpvhcrpphnastgpnwvlpm 64  
 QY 49 -----SQOVA-----L 54

Db 65 gpngkperclrfvhpnaslpndtqramepcldgwyynstkdsivtewdlvcnsnklkem 124  
 QY 55 LTVSVFVGMSSSTLWGNISDOYGRKTGLKISVLWTLTYGILSAFAPVYSWILVRLGLVG 114  
 Db 125 aqsfmagilvgilvgldisdrfgrpiltcsyllaasgsgaafstfplymvfrfclg 184  
 QY 115 FGIGGVPOS-VTLYAEFLPMKARAKCILLIEVFMAICTVFEVVLAVFVMPISLG-----WR 168  
 Db 185 fgisgtlsvlinvewpstraimstaigycytifgq-----filpglayaipqr 236  
 QY 169 WLLILSAVPLLLFAVLCFWLPESARYDVLGSGNOEKATLKRAT-----ENGAPMPLG--- 222  
 Db 237 wlqitvsiptffvflswvwtpesirwlvjsgksskalkilrrvgclqwgqgrrrkislee 296  
 QY 223 -KLIIISROEDRGKMR-----DLF-TPHFRWTTLLWFWFNSAFSYYGLVL 266  
 Db 297 lknldqkdsakanytasdlfrpmlrrmtfclsawfatgfayyslam 346

## RESULT 12

AAW49401  
 ID AAB49401 standard; Protein; 540 AA.  
 XX  
 AC AAB49401;  
 XX  
 DT 02-MAR-2001 (first entry)  
 XX  
 DE Murine organic anion transporter 6.  
 XX  
 KW Murine; organic anion transporter 6; mOATP6; cancer; inflammation;  
 KW cardiovascular disease; central nervous system disorder; kidney disease;  
 KW liver disease; autoimmune disease.  
 XX  
 OS Mus sp.  
 XX  
 PN WO200070048-A1.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 15-MAY-2000; 2000WO-US13316.  
 XX  
 PR 14-MAY-1999; 99US-0134137.  
 PR 12-MAY-2000; 2000US-0570293.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Feild J, Yue L, Ellens H;  
 XX  
 DR WPI; 2001-016235/02.  
 DR N-PSDB; AAC83979.  
 XX  
 PT Murine organic anion transporter 6 polypeptide, useful for identifying  
 PT agonists/antagonists that are useful in treatment of cancer, kidney  
 PT disease, autoimmune disease, inflammation and cardiovascular disease -  
 XX  
 PS Claim 2; Page 28-29; 32pp; English.  
 XX  
 CC The present sequence is murine organic anion transporter 6 (mOATP6).  
 CC mOATP6 protein is useful for screening compounds which inhibit or  
 CC stimulate the function of mOATP6 and also compounds that neither agonise  
 CC nor antagonise OATP6. The identified agonists and antagonists are useful  
 CC for prevention and treatment of human diseases, including cancer,  
 CC inflammation, cardiovascular disease, central nervous system disorders,  
 CC kidney diseases, liver disease and autoimmune diseases.  
 XX  
 SQ Sequence 540 AA;

Query Match 17.9%; Score 256.5; DB 22; Length 540;  
 Best Local Similarity 29.4%; Pred. No. 2e-20;  
 Matches 76; Conservative 47; Mismatches 117; Indels 23; Gaps 8;

```
QY 16 LSVLTCLAWADAMEMMILSLILAPOLHCEWRLPSSQVAL---LTSVVFVGMSSSTLWGN 72
Db 109 ltvpsqgweydrse-----fstiatewdlvcdqrglnkvtsctffigvilgavvvey 162
QY 73 ISDOYGRKTKLKSIVLWTLTYGILSAFAPVYSWILVRLGLVGFGGVPOSV-TLYAEFL 131
Db 163 lsdrgfrlllvaystlaiglmsaasvnyimfvttmllgsalagftiivlplewl 222
QY 132 PKARAKCILLIEVFWAIGTVEVVLAVFVMPVSLQWRLLILSAYPLLLFAVLCWLPES 191
Db 223 dvehrtvqvisttwt-ggvllltlvgyllrs--wrwlllaatlpcvpgiiswwvpes 279
QY 192 ARYDVLGSGNQERKAIATLKRKRIATPENGAPM-----PLCKLI-ISRQEDRGKMRDLF-TP 241
Db 280 arwlitqgrveaakylsacakingrpsedsalsqeaInkvitmervsqrpsyldlrts 339
QY 242 HFRWTTLLWFIWFSNAFSYGVLV 266
Db 340 qirhvsllccmmmwfgvnfsyygltl 364

RESULT 13
AAW64538
ID AAW64538 standard; Protein: 554 AA.
XX
AC AAW64538;
XX
DT 21-OCT-1998 (first entry)
XX
DE Human liver cell clone HP01293 protein.
XX
KW Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW differentiation; immune system; stimulator; suppressor; regulator;
KW hematopoiesis; activin; inhibitor; chemotactic; chemokine; receptor;
KW haemostatic; thrombolytic; ligand; anti-inflammatory; tumour.
XX
OS Homo sapiens.
XX
PN WO9821328-A2.
XX
PD 22-MAY-1998.
XX
PF 07-NOV-1997; 97WO-JP04056.
XX
PR 13-NOV-1996; 96JP-0301429.
XX
PA (PROT-) PROTEGENE INC.
PA (SAGA ) SAGAMI CHEM RES CENTRE.
XX
PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;
DR WPI: 1998-297932/26.
DR N-PSDB; AAV49558, AAV49559.
XX
PT Human protein having transmembrane domain - useful for, e.g.
research and nutrition
XX
PS Claim 1; Page 96-98; 205pp; English.
XX
CC AAW64534-W64558 represent human proteins containing a transmembrane
domain. These proteins can be used for, e.g. research and nutrition, and
may have cytokine and cell proliferation/differentiation, immune
stimulating/suppressing, haematopoiesis regulating, tissue growth,
CC activin/inhibin, chemotactic/chemokine, haemostatic and thrombolytic,
XX receptor/ligand, anti-inflammatory or tumour inhibition activity.
XX
SQ Sequence 554 AA;
```

Query Match 17.8%; Score 255; DB 19; Length 554;  
Best Local Similarity 31.1%; Pred. No. 3, 1e-20;  
Matches 70; Conservative 42; Mismatches 103; Indels 10; Gaps 5;

```
QY 49 SNOVALLTSVVFVGMSSSTLWGNISDOYGRKTKLKSIVLWTLTYGILSAFAPVYSWILV 108
Db 145 swkidlfscinagffgslgvgyfadrgrklcllgtvlvnavsgvlmafspnymsmll 204
QY 109 LRLGVGF-GIGVPOSVTLYAEFLPMKARAKCILLIEVFWAIGTVEVVLAVFVMPVSLCW 167
Db 205 frllqglvskgnmagytlitfvgsgsrtrvaimyqmaftvglvaltgla-yalph--w 261
QY 168 RWLLTILSAVPLLLFAVLCFWLPESARYDVLGSGNQERKAIATLKRKRIATPENGAPMPLGKLIIS 227
Db 262 rwlqvavslptflilywcvpsrwlisqkrnteaikimdhiaqknkplpadikmls 321
QY 228 ROED-----RGKMRDLF-TPHFRWTTLLWFIWFSNAFSYGVLV 266
Db 322 leedvteklspsfadlfrtrprkrtrfilmylwfldsvlyqglil 366

RESULT 14
AAW67617
ID AAW67617 standard; Protein: 399 AA.
XX
AC AAW67617;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:416.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-Q141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
```

```

DR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042080.
XX
XX (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-071486/08.
XX N-PSDB; AAF67950.
XX
XX Corynebacterium glutamicum nucleic acids encoding membrane construction
XX and membrane transport proteins or their portions, useful for typing or
XX identifying C. glutamicum or related bacteria, and as markers for
XX transformation -
XX
XX Claim 20; Page 759-760; 1119pp; English.
XX
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
XX construction and membrane transport (MCT) proteins given in AAF76510 to
XX AAF76847. The MCT nucleic acids and proteins are useful in the
XX identification of microorganisms which can be used to produce fine
XX chemicals, for modulating fine chemical production in C. glutamicum or
XX related bacteria (e.g. Brevibacterium lactofermentum), the typing or
XX identification of C. glutamicum or related bacteria, as reference points
XX for mapping C. glutamicum genome, and as markers for transformation.
XX AAF68082 and AAF68082 represent sequencing primers which are used in an
XX example from the present invention.
XX
XX Sequence 399 AA;
XX
XX Query Match 17.2%; Score 246.5; DB 22; Length 399;
XX Best Local Similarity 27.4%; Pred. No. 1.8e-19;
XX Matches 71; Conservative 47; Mismatches 126; Indels 15; Gaps 7;
XX
XX Qy 16 LSVLTGLAAMA--DAMEMMILSLAPOL--HCEWRLPSQVVALITSVFVGMSSSTLW 70
XX Db 7 vgtvllwfaivdvgfvlvgatipmledpawdlcagatqistlgvmtigaltl 66
XX
XX Qy 71 GNISDQYGRKTKLKSIVLWTLTYGILSAFAPVSMILVRLGLVGFGIQG-VPOSVTLVAAE 129
XX Db 67 gftldrlgrrrvmfsvavsvftlllafttnvqlfslwrlagvlggaipaiamvte 126
XX
XX Qy 130 FLPMKARAKCLLILEVFWAIGTVEVVLAVFVMPSLGHRWLLILSAYP-LLLFAVLCPWL 188
XX Db 127 frpgtkagsattlmtgyhgvavatqfllgldgfgwhsmfigavpgllilpllyffl 186
XX
XX Qy 189 PESARYDVLNSQNBKAIATLKRRIATENGAPMLPKLIIISROEDRGK---MRDLFTPHFRW 245
XX Db 187 pespyklksgkldeaq---vaasyglsid-dlddreheeelegssisslfxpsfir 241
XX
XX Qy 246 TLLLWFIWFNSAFSYGL 264
XX Db 242 ntlaiwgtsgmgllyvgl 260
XX
XX RESULT 15
XX AAG90943
XX ID AAG90943 standard; Protein; 431 AA.
XX
XX AC AAG90943;
XX
XX DT 26-SEP-2001 (first entry)
XX
XX DE C glutamicum protein fragment SEQ ID NO: 4697.
XX
XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX OS Corynebacterium glutamicum.
XX

```

PN	EP1108790-A2.
XX	
PD	20-JUN-2001.
XX	
PX	18-DEC-2000; 2000EP-0127688..
XX	
PR	16-DEC-1999; 99JP-0377484..
PR	07-APR-2000; 2000JP-0159162..
XX	
PR	03-AUG-2000; 2000JP-0280988..
XX	
PA	(KYOW ) KYOWA HAKKO KOGYO KK.
XX	
PI	Nakagawa S., Mizoguchi H., Ando S., Hayashi M., Ochiai K., Yokoi H.;
PI	Tateishi N., Senoh A., Ikeda M., Ozaki A;
XX	
DR	WPI: 2001-376931/40.
DR	N-PSDB; AAH66162.
XX	
PT	Novel polynucleotides derived from Coryneform bacteria, for identifying
PT	mutation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
XX	
PS	Claim 17; SEQ ID NO: 4697; 245pp + Sequence Listing; English.
XX	
CC	The present invention provides a number of nucleotide and protein
CC	sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC	are useful for identifying the mutation point of a gene derived from a
CC	mutant of coryneform bacterium, measuring expression amount and
CC	analysing the expression profile or expression pattern of a gene derived
CC	from Coryneform bacterium, and identifying a homologue of a gene derived
CC	from coryneform bacterium. Coryneform bacteria are useful for producing
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC	particularly L-lysine. The present sequence is a protein described
CC	in the exemplification of the invention.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from the
XX	European Patent Office.
XX	
SQ	Sequence 431 AA:
	Query Match 17.2%; Score 246.5; DB 22; Length 431;
	Best Local Similarity 27.4%; Pred. NO. 2e-19;
	Matches 71; Conservative 47; Mismatches 126; Indels 15; Gaps 7
QY	16 LSVLTGLAWMA--DAMENMILSLAPOL--HCEWRLPSSQVALLTSVVVFVGMMSSSTLW 70
Db	13 vgtvtallwfaivldgdfvlgatipmsledpawdltagqtqstlglvmftigalti 72
QY	71 GNISDYGRKTKGISVLWTYYGILSAPFAPYSWILVRGIIVGFIGGS-VPOSVTLYAE 129
Db	73 gfltdrlgrrrmlfsvavfsvftlllafttnqvlfslwrflagvglggaalptaiamvte 132
QY	130 FLPMKARAKCILLIEVFNAIGTVFEVWLAVFVWPSLGHRWLLIILSAVP-LLLFAVLCPWL 188
Db	133 frpgtktagsastlmntgyhvgevatafiglfiidfgwhsmfiagavpglilplyffli 192
QY	189 PESARYDVLSNGOEKAATLKRIATENAGMPGLKLIISRQDRCK---MRDLFTPHFRW 245
Db	193 pespyklisgkldaqa-----vaasyglisd-ddldreheelgcsslslfkpsfir 247
QY	246 TLLLWTFWFSNAFSY YGL 264
Db	248 ntlaiwqtsfmqlilvyql 266

Search completed: March 13, 2002, 12:41:07  
Job time: 54 sec

GenCore version 4.5  
Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:43:20 ; Search time 21.65 Seconds  
(without alignments)  
283.760 Million cell updates/sec

Title: US-09-911-667A-4  
Perfect score: 1429  
Sequence: 1 EDAVEAIGFGKFWKLSVLT.....WFSNAFSYGLVLLTTELQ 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	281	19.7	556	3	US-08-501-572-1
2	281	19.7	556	3	US-09-040-444-1
3	271	19.0	555	3	US-08-501-572-3
4	271	19.0	555	3	US-09-040-444-3
5	270.5	18.9	537	2	US-08-647-397-2
6	255	17.8	553	3	US-08-501-572-2
7	255	17.8	553	3	US-09-040-444-2
8	154.5	10.8	584	2	US-08-928-692-13
9	145.5	10.2	524	2	US-08-928-692-12
10	138	9.7	493	4	US-09-031-392-10
11	138	9.7	493	4	US-09-299-549-10
12	138	9.7	500	2	US-09-031-392-7
13	138	9.7	500	4	US-09-299-549-7
14	131.5	9.2	509	2	US-09-031-392-6
15	131.5	9.2	509	4	US-09-299-549-6
16	130.5	9.1	534	2	US-09-031-392-4
17	130.5	9.1	534	4	US-09-299-549-4
18	129.5	9.1	520	4	US-08-964-127-2
19	128	9.0	492	2	US-08-355-844-3
20	128	9.0	492	5	PCT-US95-16126-3
21	124	8.7	286	4	US-08-964-127-4
22	118.5	8.3	488	2	US-08-928-692-11
23	106	7.4	494	2	US-09-031-392-5
24	106	7.4	494	4	US-09-299-549-5
25	105	7.3	387	2	US-08-872-302-5
26	104.5	7.3	396	2	US-08-850-880-4
27	104.5	7.3	396	2	US-08-944-916-4

28	104.5	7.3	396	2	US-08-814-877-4	Sequence 4, Appli
29	104.5	7.3	457	2	US-08-882-704A-6	Sequence 6, Appli
30	103.5	7.2	455	1	US-08-035-928-2	Sequence 2, Appli
31	103	7.2	488	2	US-08-928-692-10	Sequence 10, Appli
32	98	6.9	383	2	US-09-031-392-3	Sequence 3, Appli
33	98	6.9	383	4	US-09-299-549-3	Sequence 3, Appli
34	96	6.7	514	1	US-08-063-552-13	Sequence 13, Appli
35	96	6.7	514	5	PCT-US93-05704-13	Sequence 13, Appli
36	95.5	6.7	563	2	US-09-031-392-2	Sequence 2, Appli
37	95.5	6.7	563	4	US-09-299-549-2	Sequence 2, Appli
38	94	6.6	521	1	US-08-063-552-2	Sequence 2, Appli
39	94	6.6	521	5	PCT-US93-05704-2	Sequence 2, Appli
40	93	6.5	515	1	US-08-063-552-4	Sequence 4, Appli
41	93	6.5	515	5	PCT-US93-05704-4	Sequence 4, Appli
42	93	6.5	528	2	US-08-403-852D-21	Sequence 21, Appli
43	93	6.5	528	3	US-08-510-646B-22	Sequence 22, Appli
44	93	6.5	528	4	US-09-231-818-21	Sequence 21, Appli
45	90.5	6.3	423	2	US-08-494-907-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1  
US-08-501-572-1  
; Sequence 1, Application US/08501572  
; Patent No. 6063623  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; APPLICANT: Gorboulev, Valentin  
; TITLE OF INVENTION: Transport protein Which Effects The  
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,  
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/501,572  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toohey, Kimberlin M  
; REGISTRATION NUMBER: 35,391  
; REFERENCE/DOCKET NUMBER: 02481.1453-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 556 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-501-572-1

Query Match 19.7%; Score 281; DB 3; Length 556;  
Best Local Similarity 32.0%; Pred. No. 1.8e-22;  
Matches 72; Conservative 49; Mismatches 94; Indels 10; Gaps 5;

QY 49 SQVALLTSVVFVGMSSSTLWGNISDQYGRKTKLSVLWTLTYGILSAFAPYISWILY 108

Db 146 ANKVDLFQSCVNLGFFGLSLVVGYYADRGKRLCLLVTTLVTSVSGVLTAFAVDYTSMLL 205  
QY 109 LRGLVGF-GIGVPOSVTLYAEFLPMKARAKICILLIEVFWAIGTVFVVLAVFVMPSLGW 167  
Db 206 FRLLQGVSKGSWVSGYTLITEFVGSYRRRTTALYQMAFTVGLVGLAGVA-YAIPD--W 262  
QY 168 RWLLILSAVPLLLFAVLCFWLPESARYDVLSGNOEKAIAATLKRIATENGAMPGLKLIIS 227  
Db 263 RWQLAVSLPTFLFLLYWVFPESRWLLSQKRTTRAVRIMEQIAOKNGKVPADLKMILC 322  
QY 228 ROEDRGKMR-----DLF-TPHFRWTTLLLWFIWFSNAFSYYGLVL 266  
Db 323 LEEDASEKRSPSFADLFRTPNLKRHTVILMYLWFSCAVLYOGLIM 367

RESULT 2  
US-09-040-444-1  
; Sequence 1, Application US/09040444  
; Patent No. 6063766  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; APPLICANT: Gorboulev, Valentin  
; TITLE OF INVENTION: Transport protein Which Effects The  
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,  
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner, L.L.P.  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/040.444  
; FILING DATE: March 18, 1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: O'Connor, Steven P  
; REGISTRATION NUMBER: 41,225  
; REFERENCE/DOCKET NUMBER: 2481.1453-01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 556 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-040-444-1

Query Match 19.7%; Score 281; DB 3; Length 556;  
Best Local Similarity 32.0%; Pred. NO. 1.8e-22;  
Matches 72; Conservative 49; Mismatches 94; Indels 10; Gaps 5;  
QY 49 SMOVALLTSVVFVGMSSSTLWGNISDOYGRKTGLKISVLWTLTYLILSAFAPVSWILY 108  
Db 146 ANKVDLFQSCVNLGFFGLSLVVGYYADRGKRLCLLVTTLVTSVSGVLTAFAVDYTSMLL 205  
QY 109 LRGLVGF-GIGVPOSVTLYAEFLPMKARAKICILLIEVFWAIGTVFVVLAVFVMPSLGW 167  
Db 206 FRLLQGVSKGSWVSGYTLITEFVGSYRRRTTALYQMAFTVGLVGLAGVA-YAIPD--W 262

QY 168 RWLLILSAVPLLLFAVLCFWLPESARYDVLSGNOEKAIAATLKRIATENGAMPGLKLIIS 227  
Db 263 RWQLAVSLPTFLFLLYWVFPESRWLLSQKRTTRAVRIMEQIAOKNGKVPADLKMILC 322  
QY 228 ROEDRGKMR-----DLF-TPHFRWTTLLLWFIWFSNAFSYYGLVL 266  
Db 323 LEEDASEKRSPSFADLFRTPNLKRHTVILMYLWFSCAVLYOGLIM 367

RESULT 3  
US-08-501-572-3  
; Sequence 3, Application US/08501572  
; Patent No. 6063623  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; APPLICANT: Gorboulev, Valentin  
; TITLE OF INVENTION: Transport protein Which Effects The  
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,  
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/501,572  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toohey, Kimberlin M  
; REGISTRATION NUMBER: 35,391  
; REFERENCE/DOCKET NUMBER: 02481.1453-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 555 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-501-572-3

Query Match 19.0%; Score 271; DB 3; Length 555;  
Best Local Similarity 31.9%; Pred. NO. 2.2e-21;  
Matches 74; Conservative 43; Mismatches 91; Indels 24; Gaps 8;

QY 49 SMOVALLTSVVFVGMSSSTLWGNISDOYGRKTGLKISVLWTLTYLILSAFAPVSWILY 108  
Db 146 SMLDLFQSSVNVGFFIGNSIGYIADRFGRKLCCLTVLINAAGVLMASPTYTWMLI 205  
QY 109 LR---GLV---GFGIGVPOSVTLYAEFLPMKARAKICILLIEVFWAIGTVFVVLAVFV 160  
Db 206 FRLLQGLVSKAGWLIGYI-----LITEFVGGYRRRTVGVYVAVTVGL---LVLAGVAY 257  
QY 161 VPSLQGRWLLILSAVPLLLFAVLCFWLPESARYDVLSGNOEKAIAATLKRIATENGAMP 220  
Db 258 ALPH--WRWLQFTVALPNFFLLYYWCIPESPRWLISQNKNAEMRIKHIKNGKSLP 315  
QY 221 LCKLIIHQEDRGK-----MRDLF-TPHFRWTTLLLWFIWFSNAFSYYGLVL 266  
Db 316 ASLQRLRLEETGKLNPSFLDLVRTPQIRKHTMILMYNFTSSVLYOGLIM 367



APPLICANT: Beier, David R.  
APPLICANT: Brady, Kevin P.  
TITLE OF INVENTION: OSTEOCLAST TRANSPORTER  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/647,397  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Gates, Edward R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: B0801/7048  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 537 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-647-397-2

Query Match 18.9%; Score 270.5; DB 2; Length 537;  
Best Local Similarity 24.6%; Pred. No. 2.4e-21;  
Matches 86; Conservative 51; Mismatches 120; Indels 93; Gaps 11;  
Qy 2 DAVEAIG-FGKFWKLSVLTGLAMADAME--MMILSLAPOLICE-----WRLP- 48  
Db 5 EILDVGSNGPQYLHVTLALPILGIANHLLQIFTATPDHHCRRPPPNASLEPWLPL 64  
Qy 49 -----SNOVA-----L 54  
Db 65 GPNKPEKCLRFVHLPLNASLPNDTOGATEPCLDGHIYNSTRDTIVTEWLVGSKNKLKEM 124  
Qy 55 LTSVVFVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVLKGLVG 114  
Db 125 AQSVFMAGILVGGPVFGEISDRFGRKPILTWSYLLLAASGSAFSPSLTYVMIFRFLCG 184  
Qy 115 FGIGVPOS-VTLYAEFLPMKARAKCILLIEVFVAIGTVFEVVLAVFVMPSLG-----WR 168  
Db 185 CSISGISLSTILNVWVPTSTRAISSTTGYCYTIGO-----FILPLGAYAVPQWR 236  
Qy 169 WLLILSAVPLLFAVLCFWLPESARYDVLGNGQEKATATLKRIAT-----ENGAPMPLGK 223  
Db 237 WLQLSVAFAFFISLLSWVPESIRWLVLGSKFSRAKLTQRTVATFNGKKEGKLTVEE 296  
Qy 224 LIISQED-----RGKWRDLF-TPHFWTTLWLFVFNFAFYSYGLVL 266  
Db 237 LKFNLOKQDITSAKVYGLSDLFRVLSILRVTFCLSLAWFATGATFYISLAM 346

RESULT 6  
US-08-501-572-2  
; Sequence 2, Application US/08501572  
; Patent No. 6063623  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; APPLICANT: Gorboulev, Valentin

US-09-040-444-3  
; Sequence 3, Application US/09040444  
; Patent No. 6063766  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Grundeman, Dirk  
; APPLICANT: Gorboulev, Valentin  
; TITLE OF INVENTION: Transport protein Which Effects The  
; TITLE OF INVENTION: Transport of Cationic Xenobiotics and/or Pharmaceuticals,  
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,444  
FILING DATE: March 18, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Connor, Steven P  
REGISTRATION NUMBER: 41,225  
REFERENCE/DOCKET NUMBER: 2481.1453-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)408-4000  
TELEFAX: (202)408-4400  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 555 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-040-444-3

Query Match 19.0%; Score 271; DB 3; Length 555;  
Best Local Similarity 31.9%; Pred. No. 2.2e-21;  
Matches 74; Conservative 43; Mismatches 91; Indels 24; Gaps 8;  
Qy 49 SNOVALLTSVFPVGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGILSAFAPVYSWILV 108  
Db 146 SMDLFOSSVNVGFFIGMSIGYIADRFGRKLCCLTTVLNAAAGVLMASPTVWMLI 205  
Qy 109 LR---GLV---GFGIGVPOSVTLYAEFLPMKARAKCILLIEVFVAIGTVFEVILA--VF 160  
Db 206 FRLOGLVSKAGWLICYI-----LITEFVGGRYRTVGIFYQAVTVGL--LVLGWAY 257  
Qy 161 VMSLGRWLLLSAVPLLLFAVLCFWLPESARYDVLGNGQEKATATLKRIATENGAPMP 220  
Db 258 ALPH--WRWLOFTVALPFFLLYYWCIPESPRWLISONKNAEAMRIKTKAKNGKSLP 315  
Qy 221 LCKLIISROEDRGK-----MRDLF-TPHFWTTLWLFVFNFAFYSYGLVL 266  
Db 316 ASLQRLLEETGKGLNLSFLDLVTPQIRKHTMLMYNFTSSVLYOGLIM 367

RESULT 5  
US-08-647-397-2  
; Sequence 2, Application US/08647397  
; Patent No. 5972702  
; GENERAL INFORMATION:



```

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0123
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5958727e
; US-08-928-692-13

Query Match 10.8%; Score 154.5; DB 2; Length 584;
Best Local Similarity 28.4%; Pred. No. 1.3e-08;
Matches 58; Conservative 36; Mismatches 93; Indels 17; Gaps 8;

QY 54 LLTSVVFVGMSSSTLWGNISDQYGRKTKISVLWTLTYGILSAFAVPSWILVLRGLV 113
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 IVTAATSLGALITSIFAGTAADIFGRKRLGMSNLMEVIGAILLOVSAHTFQMAV^RLIM 189
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 114 GFGIG-GVPSQVTLVAFELPMKARAKILLIEVFWAIGTVFVVL-----AVFVMPSLGWR 168
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 190 GFGVGIGSLIAPLFISEIAPKMIIRGR-LTVINSILWLTGG--QLVAYCGGAGLNYYNNGWR 246
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 169 WLLTSLVAPLL-FAVLCFWLPESARYDVLGSGNOEKAIATLKRATENGAPM---PLGKL 224
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 ILVGLSLIPTAVQFTCLCF-LPDTPRYYVMKGLARATEVLRKSYTDTSEIIERKVEEL 305
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 225 IISRQEDRGKMRDLFTPHFRWTL 248
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 VTLNOSIPGKN---VPEKVWNTI 325
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-08-928-692-12
; Sequence 12, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; CORRESPONDENCE ADDRESSES:
; NUMBER OF SEQUENCES: 80
; ADDRESS: No. 5958727o No. 5958727disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0123
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 524 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5958727e
; US-08-928-692-12

Query Match 10.2%; Score 145.5; DB 2; Length 524;
Best Local Similarity 25.4%; Pred. No. 1.1e-07;
Matches 62; Conservative 45; Mismatches 104; Indels 33; Gaps 9;

QY 53 ALLTSVVFVGMSSSTLWGNISDQYGRKTKISVLWTLTYGILSAFA---PVYSWILVLL 109
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 SLSVSSFAVGCMTASFFCGWLGDTLGRKAMLVANILSVGALLMGFSKLGPSHILLIAG 157
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 110 RGLNGFGIGGVQSVTLV-AEFLPMKARAKILLIEVFWAIGTVFVVLAVFVMPFS---- 164
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 RSLQGLVGLSLGSLVPMYIGETAPALRG-----ALGTFHQLAIVTGILISQIIG 207
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 165 ---LG---WRLLILLSAVPLLLFAVLCFWLPESARYDVLGSGNOE-KAIAITLKRATEN 215
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 LEFILGNVDLWHLLGSGVRAILOSLULFCPEPSRPLYIKLDEEVAKQSLKRUGYD 267
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 216 GAPPLGKLLIISROEDRGKMR---DLFT-PHFRWTTLLLMFIMFSNAFS-YYGLVLLTT 269
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 DVTRKDINEMRKEEREASSEKQVSIILQFTNSSYRQPILVALMLHVAQQFSGINGIFYYST 327
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 270 ELFQ 273
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 SIFQ 331
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
US-09-031-392-10
; Sequence 10, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
```

```

US-09-299-549-10
Query Match          9.7%; Score 138; DB 4; Length 493;
Best Local Similarity 25.0%; Pred. No. 6.7e-07;
Matches 59; Conservative 43; Mismatches 88; Indels 46; Gaps 9;

QY 4 VEATGFGKFWKLSVLTGLAWMADAMEMMILSIAPQLHCEWRPLPS-----WQVALLT 56
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 20 I AALGSFQYGNLGV-----NAPQKVIEAFYETWLGXGEXPVPTLTLLW--SLSV 70
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 57 SVVPVGMSSSTLWGNISDOYGRKTGLKISVLWTLTYGILSAFAPV---YSWILVLRGLV 113
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 71 SIFAVGCMGSFLVXIGNRLGRKXAMLVNVLAIAGLLMGLAKXAXSFEMLIIGRFII 130
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 114 GFGTGGVQPSYTL-ABFLPMKARAKCITLLIEFWAIGTVEF--VVLAVFVMPISIG---- 166
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 131 GLYGLSSGVPMVVGIEISPTALRG-----ALGTINOLGIVIGILIAQVLGLDLSL 180
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 167 -----WRWLLISALVPLLFLAVLCFWLPESARYDVLNQGJ-----KAIAITLKRIA 212
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 181 LGNESLPLLLGLTGVPAIIQLLLLPCCPESPRYLLINKNEARAKKALQRLRCTA 236
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 12
US-09-031-392-7
: Sequence 7, Application US/09031392
: Patent No. 5942398
: GENERAL INFORMATION:
: APPLICANT: Tartaglia, Louis A.
: APPLICANT: Weng, Xun
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES
: TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows95
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/031,392
: FILING DATE: 26-FEB-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Meiklejohn, Ph.D., Anita L.
: REGISTRATION NUMBER: 35,283
: REFERENCE/DOCKET NUMBER: 07334/072001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/542-5070
: TELEFAX: 617/542-8906
: TELEX: 200154
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 500 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-09-031-392-7

Query Match          9.7%; Score 138; DB 2; Length 500;
Best Local Similarity 26.8%; Pred. No. 6.8e-07;
Matches 64; Conservative 35; Mismatches 102; Indels 40; Gaps 9;

QY 50 WQVALLTSVVVFGVMSSSTLWGNISDOYGRKTKLGKISVLWTLTYYGIL---SAFAPVYSWI 106
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 71 WSVTV--SMPEPGGFTGSIAGVPLVKNKFGKGAIFNNIFSTVDAIILMGC SRVATSFELI 128
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |


```

QY 107 LVLRLGVGFGIGVQSVTLV-AEFLPMKARAKCILLIEVFVAIGTVEVVLAVFVMPSL 165  
DB 129 IISRLLVGICAGVSSNVVPMYLGELAPKNLRCALGVVPLQFIVTGI---LVAQIFGLRNL 185  
QY 166 -----GWRMLLILSAVPLLLFAVLCFVLPESARYDVLGSGNOEKATATLKRIATENCAMP 220  
DB 186 LANVDGWPILLGLTGVPAAALQLLLLPPFPSPRYLLIQQKDE--AAAKALQTLRGWDSV 243  
QY 221 LKGLIISROEDRG-----KMRDLFTPIHFRWTTLLLWFIWFS-----NAFSY 261  
DB 244 DREVAEIRQDEAEAKAGFISVLKLFMRSL-----RMQLLSIIVLMGGQQLSGVNAIYV 298  
QY 262 Y 262  
DB 299 Y 299

## RESULT 13

US-09-299-549-7  
; Sequence 7, Application US/09299549  
; Patent No. 6136547  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Weng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/299,549  
; FILING DATE: 26-APR-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/031,392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 500 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-299-549-7

Query Match 9.7%; Score 138; DB 4; Length 500;  
Best Local Similarity 26.6%; Pred. No. 6.8e-07;  
Matches 64; Conservative 35; Mismatches 102; Indels 40; Gaps 9;  
QY 50 WQVALLTSVVFVGMSSSTLMGNISDQYGRKTGLKISVLTLYLGIL---SAFAPVYSWI 106  
DB 71 WSVTV--SMFPFGFGLGLVGNKFRKGLALLFNFIPIVAILMGCSSRVATSELI 128  
QY 107 LVLRLGVGFGIGVQSVTLV-AEFLPMKARAKCILLIEVFVAIGTVEVVLAVFVMPSL 165

DB 129 IISRLLVGICAGVSSNVVPMYLGELAPKNLRCALGVVPLQFIVTGI---LVAQIFGLRNL 185  
QY 166 -----GWRMLLILSAVPLLLFAVLCFVLPESARYDVLGSGNOEKATATLKRIATENCAMP 220  
DB 186 LANVDGWPILLGLTGVPAAALQLLLLPPFPSPRYLLIQQKDE--AAAKALQTLRGWDSV 243  
QY 221 LKGLIISROEDRG-----KMRDLFTPIHFRWTTLLLWFIWFS-----NAFSY 261  
DB 244 DREVAEIRQDEAEAKAGFISVLKLFMRSL-----RMQLLSIIVLMGGQQLSGVNAIYV 298  
QY 262 Y 262  
DB 299 Y 299

## RESULT 14

US-09-031-392-6  
; Sequence 6, Application US/09031392  
; Patent No. 5942398  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Weng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/031,392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 509 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-031-392-6

Query Match 9.2%; Score 131.5; DB 2; Length 509;  
Best Local Similarity 28.8%; Pred. No. 3.6e-06;  
Matches 61; Conservative 27; Mismatches 77; Indels 47; Gaps 8;

QY 53 ALLTSVVFVGMSSSTLMGNISDQYGRKTGLKISVLTLYLGILSAFAPVYSWILV 108  
DB 82 ALSVAIFSVGGMISFSLIGIISQWLGRKRAMLVNNVLAVGLSGSLMGLANA-AASYEMLIL 140  
QY 109 LRGVLGVGIGVQSVTLV-AEFLPMKARAKCILLIEVFVAIGTVEVVLAVFVMPSL 165  
DB 141 GRFLIGAYSGLTSLGLVPMYVGEIAPTHLRG-----ALGTINQIAIVIGILIAQVL 190  
QY 166 G-----WRMLLILSAVPLLLFAVLCFVLPESARYDVLGSGNOE-KATATLKRIA-- 212  
DB 191 GLESLLTASLWPLLLGLTGLTLPALLQLVLLPFCPSPRYLIIONLEGSPARKSLKRLTGW 250

Qy	213	-----TENGAMPLGLKIISR	228
		:   :	
Db	251	ADVSGVLAELKDEKKLERERPLSLQLLGS	282

```

US-09-299-549-6
; Sequence 6, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/299,549  
FILING DATE: 26-APR-1999

PRIOR APPLICATION DATA: 09/031,392  
 APPLICATION NUMBER: 09/031,392  
 FILING DATE: 26-FEB-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Melkiejohn, Ph.D., Anita L.  
 REGISTRATION NUMBER: 35,283  
 REFERENCE/DOCKET NUMBER: 07334/072002  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 TELEX: 200154

```

; INDEX: 200134
; INFORMATION FOR SEQ ID NO: 6:
;     LENGTH: 509 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-09-299-549-6

```

```

Query Match.          9.2%; Score 131.5; DB 4; Length 509;
Best Local Similarity 28.8%; Pred. No. 3.6e-06;
Matches 61; Conservative 27; Mismatches 77; Indels 47; Gaps 8;

QY 53 ALLTSVVFVCGMSSSTLWGNISDQYGRKTKLKI-----SVLWTLTYGILSAFAPVYSILV 108
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 82 ALSVAIFSVGGMISFFLIGIISQWLGKRRKRLMNVNVLAVGGSLMGLANA--AASYEMI IL 140
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

QY 109 LRGVLGFGIGGVQSVTVLY--AEFLPMKARAKCILLIEVFVAIGTV--FEVVLVAFVMPSL 165
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 141 GRFLIGAYSGLTSGLVPYMYGEIAPHLRG-----ALGTNLQLAIVIGILIAQVL 190

QY 166 G-----WRWLLILSAVPILLFAVLCFWMLPESARYDVLGSGNOE-KAIA TLKRTA-- 212
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 191 GLESLGTASLWPLLGLTVLPALLOIVLLPFCPESPRYLIYITONLEGPARKSLKRLTGW 250
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

QY 213 -----TENGAMPMLGKLLIIR 228
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 251 ADVSGVLAELKDKRKLRLERPLSLQLQIGSR 282
      ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||

```

Search completed: March 13, 2002, 12:43:21

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:41:38 ; Search time 25.88 Seconds  
(without alignments)  
803.541 Million cell updates/sec

Title: US-09-911-667A-4  
Perfect score: 1429  
Sequence: 1 EDVAEAIQFGKFWKLSVLT.....WFSNAPSYGLVLLTTELFQ 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	831.5	58.2	529	T23190	hypothetical prote
2	707	49.5	300	S15786	glucose transport
3	446.5	31.2	454	F75580	probable sugar tra
4	358.5	25.1	422	G72234	hypothetical prote
5	348	24.4	400	C69757	transporter homolo
6	342	23.9	742	A43344	synaptic vesicle p
7	341	23.9	724	I50531	transmembrane tran
8	338	23.7	742	S27263	synaptic vesicle p
9	320	22.4	455	B83213	probable MFS trans
10	320	22.4	683	S34961	synaptic vesicle p
11	319	22.3	452	A85852	probable transport
12	281	19.7	593	JC4884	organic cation tra
13	278	19.5	536	S50862	organic cation tra
14	267.5	18.7	448	G83616	4-hydroxybenzoate
15	265	18.5	446	B83033	probable MFS trans
16	264	18.5	448	H83335	probable MFS trans
17	244.5	17.1	444	B83033	probable MFS trans
18	242.5	17.0	451	S66008	transport protein
19	237.5	16.6	401	E69501	sugar transporter
20	231	16.2	557	JW0089	organic cation tra
21	222	15.5	423	S74046	probable sugar tra
22	215	15.0	557	JE0346	high-affinity carn
23	213	14.9	576	T22509	hypothetical prote
24	210	14.7	450	C83394	probable MFS trans
25	208	14.6	457	E70070	metabolite transpo
26	205.5	14.4	464	F69587	L-arabinose transp
27	202.5	14.4	469	G65058	hypothetical prote
28	202.5	14.2	401	E85927	partial probable t
29	202.5	14.2	473	G69789	sugar transporter

30	200	14.0	522	2	S12042	glucose transport
31	200	14.0	522	2	E86246	glucose transporte
32	199	13.9	419	2	E69888	metabolite transpo
33	198.5	13.9	423	2	A83383	probable MFS trans
34	198	13.9	422	2	C70518	probable nanT prot
35	197	13.8	523	2	S25015	monosaccharide tra
36	194	13.6	409	2	T47026	hypothetical prote
37	193	13.5	443	2	E64725	yaaU protein - Esc
38	193	13.5	557	2	T38125	myo-inositol trans
39	192	13.4	443	2	H85485	probable transport
40	191.5	13.4	407	2	C64167	hypothetical prote
41	191	13.4	508	2	T05156	probable glucose t
42	186	13.0	418	2	A64763	probable transport
43	186	13.0	418	2	D85530	probable transport
44	186	13.0	516	2	T12199	monosaccharid tran
45	184.5	12.9	575	2	T43400	myo-inositol trans

ALIGNMENTS

RESULT 1  
T23190  
hypothetical protein ZK637.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Jun-2000  
C:Accession: T23190  
R:Craxton, M.  
submitted to the EMBL Data Library, April 1993  
A:Reference number: Z19704  
A:Accession: T23190  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-529 <WIL>  
A:Cross-references: EMBL:Z22175; PIDN:CAA80131.1; GSPDB:GN00021; CESP:ZK637.1  
C:Genetics:  
A:Gene: CESP:ZK637.1  
A:Map position: 3  
A:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3; 422/2; 472/3  
C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match	58.2%	Score	831.5;	DB	2;	Length	529;
Best Local Similarity	58.0%	Pred. No.	3.8e-64;				
Matches	160;	Conservative	43;	Mismatches	70;	Indels	3;
Gaps	1;						
QY	1	EDAVEAIGFGKFWKLSVLTGLAWADAMEMMLISILAPQLHCENRIPSMOVALTSVVF	60				
Db	70	DEAVEALGFRGKLSVLTGLAWADAMEMMLISILAPQLHCENRIPSMOVALTSVVF	129				
QY	61	VGMSSSTLWGNISDQYGRKTKISVLTGLAWADAMEMMLISILAPQLHCENRIPSMOVALTSVVF	120				
Db	130	SGMMLSTFWGKICDFGRKGLTSTLVACIMVSGMSPHYVLLFFRGLTGTGIGV	189				
QY	121	QSVTLVAFPLPMKARAKILLIEVFVFAIGTVFVAVVMPVSLGWRMLLSAVPLLL	180				
Db	190	QSVTLVAFPLPMKARAKILLIEVFVFAIGTVFVAVVMPVSLGWRMLLSAVPLLL	249				
QY	181	FAVLCFWLPESARYDVLVSGNOEKATATKRIATENGAPMLGLKLIISRO---	237				
Db	250	FAVASFWLPESARYDVLVSGNOEKATATKRIATENGAPMLGLKLIISRO---	309				
QY	238	LFTPHFRWTTLLLMFIFNSAFSYGLVLLTTELFQ	273				
Db	310	LLSPDLRRTILLWCIIWATAFSYGYMVLFTTVLFQ	345				

RESULT 2  
S15786  
glucose transport protein homolog - Caenorhabditis elegans (fragment)  
C:Species: Caenorhabditis elegans  
C:Date: 06-Jan-1995 #sequence\_revision 05-May-2000 #text\_change 05-May-2000

C:Accession: S15786  
R:Craxton, M.; Ainscough, R.; Coulson, A.; Dear, S.; Du, Z.; Durbin, R.; Green, P.; Hall, Staden, R.; Sulston, J.; Thierry-Mieg, J.; Thomas, K.; Waterston, R.; Wilson, R.  
submitted to the EMBL Data Library, May 1991  
A:Reference number: S15786

A:Accession: S15786  
A:Molecule type: DNA  
A:Residues: 1-300 <CRA>  
A:Cross-references: EMBL:Z11115  
C:Genetics:  
A:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3  
C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match 49.5%; Score 707; DB 2; Length 300;  
Best Local Similarity 59.0%; Pred. No. 1.1e-53;  
Matches 134; Conservative 37; Mismatches 56; Indels 0; Gaps 0;

QY 1 EDVAEATGCGKFWKLSVLTGLAWMADAMEMMILSLILAPOLHCEWRLPSPVOALLTSVVF 60  
DB 70 DEAVEALGFRGFKUSILTGLMAWADAMEMMILSLILPALACEWGISVVOQALVTCVF 129  
QY 61 VGMSSSTLWGNISDOYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGGV 120  
DB 130 SGMLSSSTFWGKICDRGRRKGLTSTFLVACIMGVISGMSPHFVYVLLFPRLTGFIGGV 189  
QY 121 POSVTLYAEPLPKAKAKCILLIEVFWAICTVFEVVLAVFVMPSLGHRWLLILSAVPLLI 180  
DB 190 POSVTLYAEPLTAQRAKCVLVIESFWAIGAVEALLAYFVMSFGWRALMFLSSILPLGI 249  
QY 181 FAVLCFWLPESARYDVLSGNQEKAIATLKRATENGAPMPGLKLIIS 227  
DB 250 FAVASFWLPESARFDMASGHPERALETLQAAARNRVQLPTGRLVSS 296

RESULT 3  
F75580  
probable sugar transporter - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: F75580  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896  
A:Accession: F75580  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-454 <WHI>  
A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12486.1; PID:g646070  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0271  
A:Map position: 2  
C:Superfamily: yaaU protein

Query Match 31.2%; Score 446.5; DB 2; Length 454;  
Best Local Similarity 35.1%; Pred. No. 4.9e-31;  
Matches 94; Conservative 52; Mismatches 115; Indels 7; Gaps 3;

QY 4 VEAIGFGKFWKLSVLTGLAWMADAMEMMILSLILAPOLHCEWRLPSPVOA--LLTSVVFV 61  
DB 19 LDDLGLGRFQWKLKLAICGLTWAADAMEVLLMGFALPGISAFAELPKGSPAATMLLTATPA 78  
QY 62 GMMSSSTLWGNISDOYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGGV-V 120  
DB 79 GMLFGAWFWGLADRVGRRSVFVTLTVALGVVFGLAGALAPTLWLLVARELTGFAIGGTL 138  
QY 121 POSVTLYAEPLPKAKAKCILLIEVFWAICTVFEVVLAVFV----MPSLGRWLLILSAV 176

DB 139 PVDYSMAAEFVPTANRGRFLVYLESEFVAVCTVVVAALAWWVSTAFAPAEGRWLLGLAAL 198  
QY 177 PLLLFAVLCFWLPESARYDVLSONQEKAIAITLKRATENGAPMPGLKLIISRQEDRGMR 236  
DB 199 PGLVGLIARIGIPDSRSLARGEQAARAAOKVAQANGCTLPAAPLAHPPEOPPRVSPA 258  
QY 237 DLFTPHFRWTTLLLWFIWFSNAFYYGL 264  
DB 259 QLFGRGLARTRPLMTVMTWFGLSLGYGJ 286

RESULT 4  
G72234  
hypothetical protein TM1603 - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: G72234  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Hatt, D.H.; H. Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome A:Reference number: A72200; MUID:99287316  
A:Accession: G72234  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-422 <ARN>  
A:Cross-references: GB:AE001804; GB:AE000512; NID:g4982160; PIDN:AAD36670.1; PID:g4982160  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1603  
C:Superfamily: hypothetical protein H11104

Query Match 25.1%; Score 358.5; DB 2; Length 422;  
Best Local Similarity 32.0%; Pred. No. 1.7e-23;  
Matches 89; Conservative 55; Mismatches 111; Indels 23; Gaps 7;

QY 1 EDVAEATGCGKFWKLSVLTGLAWMADAMEMMILSLILAPOLHCEWRLPSPVOALLTSVVF 60  
DB 4 DEIVEKYVDKTRORRFLILTSIAMFDAGVLLSFVLPVIVKEWNLTSOGATIASATF 63  
QY 61 VGMSSSTLWGNISDOYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGGV 120  
DB 64 LGMLFGALSFGFVADLLGRKVSNLFFIVITITFLSGSSFPETLLVRLGLSGFGYGL 123  
QY 121 POSVTLY-AEFLPKAKAKCILLIEVFWAICTVFEVVLAVFVMPSLGHRWLLILSAVPL 179  
DB 124 MPSFNAYLAETSTRLRGYLVILLESSWAVGSILIGLFAVNLPLN--WRWVWFISIG-Y 180  
QY 180 LFAVLCFWLPESARYDVLSONQEKAIAITL-KRIATENGAP----MPLGKLIISRQEDRGK 234  
DB 181 LFPVFLRMPETPKYAFKLGKKEALERSLGRVEEVELPKKVPILALL-----KREH 235  
QY 235 MRDLTPHFRWTTLLLWFIWFSNAFYYGLVLLTTELF 272  
DB 236 LKD-----TVVIWIAFWFVSVYYALTWPAPRIF 264

RESULT 5  
C69757  
transporter homolog yceI - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: C69757  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Beck, C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; C.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, A.; Fritsch, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; G. Authors: Foulger, D.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardir, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Me



Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Poll, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serod  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A.; Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033  
A:Accession: C69757  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-400 <KUN>  
A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CA12089.1; PID:g2632581  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yceI  
C:Superfamily: hypothetical protein H1104

Query Match 24.4%; Score 348; DB 2; Length 400;  
Best Local Similarity 33.7%; Pred. No. 1.3e-22;  
Matches 87; Conservative 51; Mismatches 86; Indels 34; Gaps 6;  
Qy 13 QWKLSTGLTGLAWMADAMEMMILSLAPOLHCEWRLPSMQVALLTSVVFVGMSSSTLWGN 72  
Db 9 QKLLGVAGLWGLFDAMDVGILSIIIAALHVEWNLSPPEMKWIGSVNSIGMAAGFLGL 68  
Qy 73 ISDQGRKTLGLISVLWTLTYGI---LSAFAPVYSWILVRLGLVGFIGG-VPOSVTLYA 128  
Db 69 LADRIGRK---KVFITLLCSIGSISAFVTSLSAFILIRFVIGMGLGELPVASTLVS 125  
Qy 129 EFLPMKARAKCILLIEVFNAITGVFEVVLAVFVMPSLGWRWLLLSAVPLLLFAVLCEWL 188  
Db 126 EAVPEKRGVILLESFWMGLAAALISYFIPSGWQAALLTALTAFYALYLRSL 185  
Qy 189 PESARYDLVSGNQEKATATKRIATENCAPMLGLKLIISROEDRGKMDLFTPHFRWTL 248  
Db 186 PDSPKYESLSAK-----KRSWEN-----VKSVMARQYIR-----PTV 218  
Qy 249 LLMFIFSNFASYGLVL 266  
Db 219 MLSIWFVCFVFSYGMFL 236

RESULT 6  
A43344  
synaptic vesicle protein SV2 - rat  
N:Alternate names: transporter-like protein p87  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Sep-2000  
C:Accession: A43344; A43267; A58948  
R:Feany, M.B.; Lee, S.; Edwards, R.H.; Buckley, K.M.  
Cell 70, 861-867, 1992  
A:Title: The synaptic vesicle protein SV2 is a novel type of transmembrane transporter.  
A:Reference number: A43344; MUID:92386605  
A:Accession: A43344  
A:Molecule type: mRNA  
A:Residues: 1-742 <FEA>  
A:Cross-references: GB:L01788; NID:g207090  
A:Experimental source: pheochromocytoma cell line PC12  
A:Note: sequence extracted from NCBI backbone (NCBIN:112840, NCBI:P:112842)  
R:Baizalieh, S.M.; Peterson, K.; Shinghal, R.; Scheller, R.H.  
Science 257, 1271-1273, 1992  
A:Title: SV2, a brain synaptic vesicle protein homologous to bacterial transporters.  
A:Reference number: A43267; MUID:92390722  
A:Accession: A43267  
A:Molecule type: mRNA  
A:Residues: 1-339 'F', 341-742 <BA>  
A:Cross-references: GB:L05435; NID:g207091; PIDN:AAA42188.1; PID:g207092  
A:Experimental source: brain  
A:Note: sequence extracted from NCBI backbone (NCBI:P:112489)  
R:Gingrich, J.A.; Andersen, P.H.; Tiberi, M.; El Mestikawy, S.; Jorgensen, P.N.; Fremear

FEBS Lett. 312, 115-122, 1992  
A:Title: Identification, characterization, and molecular cloning of a novel transport  
A:Reference number: S27263; MUID:93050176  
A:Accession: A58948  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-120, 'E', 122-249, 'W', 251-684, 'P', 686-742 <GIN>  
A:Note: only differences from the bovine translation are shown in Fig. 2  
C:Superfamily: synaptic vesicle protein SV2  
C:Keywords: transport protein

Query Match 23.9%; Score 342; DB 1; Length 742;  
Best Local Similarity 27.4%; Pred. No. 8.3e-22;  
Matches 86; Conservative 55; Mismatches 125; Indels 48; Gaps 5;  
Qy 1 EDVAEIGFGKFWKLSVLTGLANMADAMEMMILSLAPOLHCEWRLPSMQVALLTSVVF 60  
Db 153 ETILRECGHGRFOWTLFVVLGLALMADGVEFVGVLPSPAEKDMCLSDSNKGMGLIYV 212  
Qy 61 VGMSSSTLWGNISDOYGRKTLGISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGG- 119  
Db 213 LGMVGAFLWGLADRLGRRCOLLISLSVNSVEAFFSSPVQCYGTFLFCRLLSGVGIGS 272  
Qy 120 VPOSVTLYAEFLPMKARAKCILLIEVFNAITGVFEVVLAVFVMPSLG----- 166  
Db 273 IPIVSYSEFLAQKRGHEHSLWLCFMWMIIGGVYAAAMAWAIIPIHYGWSFOMGSAYOFHS 332  
Qy 167 WRWLLLSAVPLLLFAVLCEWRLPSARYDLVSGNQEKATATKRIATEN---GAP---M 219  
Db 333 WRVFLVCAFPVFAIGALTTPQSPRPLENGKDEAMWVLKQVHDTNMRAGHPERVF 392  
Qy 220 PLGLKLIISRQEDR-----GKMRDLFTPHFRWTLTLLMF 252  
Db 393 SVTHIKTHQEDLIEIOGDTGTQYRGMGRALSLGGQVMGNFLSCFSPYRRTITLMMG 452  
Qy 253 IWFSNFASYGLVL 266  
Db 453 VMTMSFSYGLTV 466  
RESULT 7  
150531  
transmembrane transporter - electric ray (Discopyge ommata)  
C:Species: Discopyge ommata  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 20-Aug-1999  
C:Accession: 150531  
R:Bindra, P.S.; Knowles, R.; Buckley, K.M.  
Gene 137, 299-302, 1994  
A:Title: Conservation of the amino acid sequence of SV2, a transmembrane transporter  
A:Reference number: 150531; MUID:94131301  
A:Accession: 150531  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-724 <BIN>  
A:Cross-references: GB:L23403; NID:g472816; PIDN:AAA49235.1; PID:g472817  
C:Superfamily: synaptic vesicle protein SV2

Query Match 23.9%; Score 341; DB 2; Length 724;  
Best Local Similarity 28.3%; Pred. No. 9.8e-22;  
Matches 89; Conservative 54; Mismatches 112; Indels 60; Gaps 7;  
Qy 4 VEAIGFGKFWKLSVLTGLANMADAMEMMILSLAPOLHCEWRLPSMQVALLTSVVF 63  
Db 139 IQECHGRFQWALFLVGLSLMADGVEFVGVLPSPAEKDMCLSDSNKGMGLIYV 198  
Qy 64 MSSSTLWGNISDOYGRKTLGISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGG-VPO 122  
Db 199 MLGAFFWGLADKMGRRQTLIFCMSSINGFFAFLLSFVQGYSLFLFCRFFAGFGIGAVP 258  
Qy 123 SVTLYAEFLPMKARAKCILLIEVFNAITGVFEVVLAVFVMPSLG-----RW 169

Db	259	VFIYFAEVLAREKRGHLSWLCMFWMIGGIYASAMAWAIIIPHYGQSFSGMSAYQFHSWKV	318
QY	170	LLILSAVPLLFAVLFCWLPESARYDVLSGNQEKAIAATLKRIATEN	218
Db	319	FVVVLCALPCMSSVALTFNPESPRVLEVGKHDAAWMLKQIHDTNMRARGQPEKFTVN	378
QY	219	-MPLKGLI::ISRQEDRCMKMRDLTFPHRW	249
Db	379	RIKTPKLDELIEIQDTG-----TWYMRWFVRIKTEYGIWLTfMRCLDYPVKRNTIL	432
QY	250	LWFIWFSNAFSYYGL	264
Db	433	LIIVWTLTSPFYGYGL	447

RESULT 8  
S27263  
synaptic vesicle protein SV2 - bovine  
N:Alternate names: transporter-like protein p87  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 22-Nov-1993 #sequence\_revision 18-Aug-2000  
C:Accession: S27263  
R:Gingrich, J.A.; Andersen, P.H.; Tiberi, M.; El Mestikawy, S.; Jorgensen, P.N.; Freemaal, E.  
FEBS Lett. 312, 115-122, 1992  
A:Title: Identification, characterization, and molecular cloning of a novel transporter-  
A:Reference number: S27263; MUID:93050176  
A:Accession: S27263  
A:Molecule type: mRNA  
A:Residues: 1-742 <GIN>  
A:Cross-references: GB:S47919; NID:Q259173; PIDN:AAB24028.1; PID:q259174  
C:Superfamily: synaptic vesicle protein SV2

Query Match 23.7%; Score 338; DB 2; Length 742;  
Best Local Similarity 27.4%; Pred. No. 1.8e-21;  
Matches 86; Conservative 54; Mismatches 126; Indels 48; Gaps 5;

[illegible]

RESULT 9  
B83213  
Probable MFS transporter PA3467 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence-revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83213  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337  
A:Accession: B83213  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-455 <S>  
A:Cross-references: GB:AE004767; GB:AE004  
A:Experimental source: strain PA01  
C:Genetics  
C:Superfamily: hypothetical protein bl775

Query Match 22.4%; Score 320; DB 2; Length 455;  
Best Local Similarity 29.4%; Pred. NO. 3.9e-20;  
Matches 85; Conservative 58; Mismatches 122; Indels 24; Gaps 6;

Qy		2	DAVEIGCFQFOWKLSVLTGLANWADAMEMHLLSILAPOLHCENRPLPSXOVALLTSSVFV	61
	:	:	: : : :	
Db		10	ERLERLPSYHRVLFFVIALAFDFSMDLAMMTFLGSIKAFFGLDSQAAGLASSSF	69
	:	:	: : : :	
Qy		62	GMMSSSTLGWNISDOYGCRKTGLKIS-VLVTLYGYLSAFAPYSWTIVRLGLVGFGIG-G	119
	:	:	: : : :	
Db		70	GMVIGAALSGMLADFRCRKVPFOASIVWLGSALYSTAGDLDLS-LTFYRVLLGTGMGE	128
	:	:	: : : :	
Qy		120	VPOSVTLYAEFLPMKARAKCILLIEVFMAITGVFEVLAVFMPSLGMWRMILLISAVPLL	179
	:	:	: : : :	
Db		129	FPIAQSLSEMPASRRGKYIAlMDGFEPWLGfVAAGLSFYLLPTLPTGWRSSIFVLVALPAV	188
	:	:	: : : :	
Qy		180	LFAVLCPWLPEsarYDVLNSGNEKATLKRIATE-----NGAMPPLGKLIIISREQDR	232
	:	:	: : : :	
Db		189	FVLAIRFLIPESPWLQSQRGRQADVLRDI EARVMSI GLTELPPPL-----RPQR	242
	:	:	: : : :	
Qy		233	GKMR-----DLTFPHRTWTTLLWF IWFSNAFSYGYLVLLTTELFQ	273
	:	:	: : : :	
Db		243	PESPFP:FFSAFATLWSPAYRRRTITVMGLWFFELFI GFGYGLTSMLSALLO	291
	:	:	: : : :	

RESULT 10

S34961  
synaptic vesicle protein 2 form B - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 20-Aug-1999  
C:Accession: S34961: A47382  
R:Bajjalieh, S.M.; Peterson, K.E.  
submitted to the EMBL Data Library, February 1993  
A:Description: Brains contains two forms of synaptic vesicle protein 2.  
A:Reference number: S34961  
A:Accession: S34961  
A:Molecule type: mRNA  
A:Residues: 1-683 <BAJ>  
A:Cross-references: EMBL:L10362; NID:g207093; PIDN:AAA42189.1; PID:g207094  
R:Bajjalieh, S.M.; Peterson, K.; Linial, M.; Scheller, R.H.  
Proc. Natl. Acad. Sci. U.S.A. 90, 2150-2154, 1993  
A:Title: Brain contains two forms of synaptic vesicle protein 2.  
A:Reference number: A47382; MUID:93211920  
A:Accession: A47382  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-683 <BAJ2>  
A:Cross-references: GB:L10362; NID:g207093; PIDN:AAA42189.1; PID:g207094  
A:Experimental source: brain  
A:Note: sequence extracted from NCBI backbone (NCBIN:128278, NCBI:P:128279)  
C:Superfamily: synaptic vesicle protein SV2

```
Query Match      22.4% Score 320; DB 2; Length 683;
Best Local Similarity 27.1%; Pred. No. 6e-20;
Matches 86; Conservative 56; Mismatches 121; Indels 54; Gaps 6;
```



Best Local Similarity 32.0%; Pred. No. 2e-16;  
Matches 72; Conservative 48; Mismatches 95; Indels 10; Gaps 5;

QY 49 SMOVALLTSVVFVGMSSSTLGNISDOYGRKTKISVLWTLTYGILSAFAPVYSWILV 108  
Db 146 AKAVDLFQSCVNLGFFGLSVGYIADRFGRKCLLVTLTVTSVGLVAVAPDYTSMLL 205  
QY 109 LRGLVGF-GIGVGPQSVTLTYAEFLPMKARAKCILLIEVFAIGTVFEVVLAVFVMPISGW 167  
Db 206 FRLLQGMVSKGSWSGYTLTIFTEFVSGYRKTALLYOMAFVGLVGLAGVA-YAIPD--W 262  
QY 168 RWLLILSAVPLLLFAVLGFWLPESARYDVLSGNQEKAIATLKRIATENGAPMPGLKLIIS 227  
Db 263 RWLQLAVALSLPTFLFLYYWFVPSRWLLSQKRTTRAVRIMEQIAOKNGKVPADLKMIC 322  
QY 228 RQDRGKMR-----DLF-TPHFRWTLTLWFIWFSNAFSYGLVL 266  
Db 323 LEEDASEKRSFADLFRPTLRKHTVILMYLWFSYGLVQLIM 367  
RESULT 14  
G83616  
4-hydroxybenzoate transporter PcaK PA0235 [imported] - Pseudomonas aeruginosa (strain PA  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: G83616  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337  
A:Accession: G83616  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-448 <STO>  
A:Cross-references: GB:AE004461; GB:AE004091; NID:99946066; PIDN:AGC03624.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: pcaK; PA0235

Query Match 18.7%; Score 267.5; DB 2; Length 448;  
Best Local Similarity 29.2%; Pred. No. 1.3e-15;  
Matches 81; Conservative 54; Mismatches 125; Indels 17; Gaps 8;

QY 4 VEAIGFGKQWKLSVLTGLAMMADAMEMMILSILAPOLHCEWRLPSWQVALLTSVVFVGM 63  
Db 18 INAQPLSPYQWRIVLCLFLVFLDGLDTAAMGFIAPALTQWIDRASLGPMVMSAALIGM 77  
QY 64 MSSSTLWGNISDOYGRKTKISVLWTLTYGILSAFAPVYSWILVLRGLVGFPGIG-GVPPQ 122  
Db 78 VFGALGSGPLADRYGRKLVVAAPVFLGFLSLSASAYSTNVEQLALRFLTLGLGLGAAMPN 137  
QY 123 SVTLTYAEFLPMKARAKCILLIEVFAIGTVFEVVLAVFVMPISGWRLILTSLA-VPLL 179  
Db 138 ATLLSEYTP--ERLKSLLVTSFMFCGFLNMGACGFSVAKLIPLFGWHSLLDGLGLPLV 195  
QY 180 LFVAVLCFWLPESARYDVL-SGNQEKAIATLKRIATENGAPMPGLKLIISROEDRGKMRDL 238  
Db 196 LAVVLLFRPLPESARYLVVNRGSRVROVLAP--AQVALARSFHVPEQQTVOARVN 252  
QY 239 FTFPHFRWT-----TLLWFIWFSNAFSYGLVLLTTEL 271  
Db 253 FAVIFSGTYSAGTLLTLMLTYFMGLVIVY---LLTSWL 286

RESULT 15  
B83033  
Probable MFS transporter PA4900 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: B83033

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p  
A:Reference number: A82950; MUID:20437337  
A:Accession: B83033  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-446 <STO>  
A:Cross-references: GB:AE004903; GB:AE004091; NID:99951173; PIDN:AGC08285.1; GSPDB:B  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4900  
C:Superfamily: yaaU protein

Query Match 18.5%; Score 265; DB 2; Length 446;  
Best Local Similarity 25.8%; Pred. No. 2.1e-15;  
Matches 70; Conservative 58; Mismatches 123; Indels 20; Gaps 6;

QY 4 VEAIGFGKQWKLSVLTGLAW-----MADAMEMMILSILAPOLHCEWRLPSWQVALLTSV 58  
Db 10 IDAARFNRTNHLI-----LGMGCFIMLFDGYDMVITYGSVVPRLMQEQLSPVQAGTLGSC 64  
QY 59 VFVGMSSSTLWGNISDOYGRKTKISVLWTLTYGILSAFAPVYS---WILVLRGLVG 114  
Db 65 ALFGMLFGGTTLLAPLADRFGRRR---LVIAITTLASLAFLTGHARDPDLGACGRFFTC 120  
QY 115 FGIGG-VPOSVTLTYAEFLPMKARAKCILLIEVFAIGTVFEVVLAVFVMPISGWRLMLLIL 173  
Db 121 LALGALVPSAINLISEFAPAGRRTSLVTVMSAFYSYGAVALSALLAIAIAMPWGQSVFYV 180  
QY 174 SAVPLLLFAVLGFWLPESARYDVLSGNQEKAIATLKRIATENGAPMPGLKLIISROEDRG 233  
Db 181 AVLPLVAVPLMLRWLPESARFLELKGRRRAELDALLRKKVDDPY-RPGAERANAVAAAPSG 239  
QY 234 KMRDLFTPHFRWTLTLWFIWFSNAFSYGL 264  
Db 240 RVAQLFEGRQAVGTLLLWVAFAMCMLMSYGL 270

Search completed: March 13, 2002, 12:41:39  
Job time: 85 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:42:52 ; Search time 17.05 seconds  
(without alignments)  
587.067 Million cell updates/sec

Title: US-09-911-667A-4  
Perfect score: 1429  
Sequence: 1 EDVAEAGKGFQWKLKSLVT.....WFSNAPSYGLVLLTTELPY 273

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	831.5	58.2	529	1 Y0UL_CAEEL	P30638 caenorhabdi
2	348	24.4	400	1 YCEI_BACSU	O34691 bacillus su
3	341	23.9	742	1 SYV2_RAT	O02563 rattus norv
4	282.5	19.8	448	1 PCAK_PSEPU	O31955 pseudomonas
5	260.5	18.2	457	1 PCAK_ACICA	Q33975 acinetobact
6	242.5	17.0	451	1 YTAJ_BACSU	P37514 bacillus su
7	238	16.7	466	1 BENK_ACICA	O30513 acinetobact
8	231	16.2	557	1 OCN2_HUMAN	O76082 homo sapien
9	227.5	15.9	413	1 PHDK_ACICA	P94131 acinetobact
10	219	15.3	473	1 PHDK_NOSK	O24723 nocardioid
11	219	15.3	557	1 OCN2_MOUSE	Q32088 mus musculu
12	219	15.3	557	1 OCN2_RAT	O70594 rattus norv
13	205.5	14.4	445	1 YGCS_ECOLI	Q46909 escherichia
14	203.5	14.2	428	1 MMLH_ALCEU	O51798 alcaligenes
15	200	14.0	522	1 STPI_ARATH	P23586 arabidopsis
16	193	13.5	443	1 YAAU_ECOLI	P31679 escherichia
17	193	13.5	557	1 ITR2_SCHPO	P87110 schizosacch
18	191.5	13.4	407	1 YB04_HAEIN	K71365 haemophilus
19	186	13.0	403	1 MHPT_ECOLI	P77589 escherichia
20	184.5	12.9	575	1 ITRI_SCHPO	Q10286 schizosacch
21	184	12.9	547	1 GTRI_LEIDO	Q01440 leishmania
22	181	12.7	523	1 STC_RICCO	Q41144 ricinus com
23	178	12.5	457	1 XYL1_LACBR	O52733 lactobacill
24	177	12.4	496	1 NANT_ECOLI	P41036 escherichia
25	175.5	12.3	751	1 YLX5_CAEEL	P46501 caenorhabdi
26	171.5	12.0	461	1 CSBC_BACSU	P46333 bacillus su
27	167.5	11.7	482	1 YFIC_BACSU	P54723 bacillus su
28	166.5	11.7	468	1 GLCP_SYNY3	P15729 synecocyst
29	162	11.3	540	1 HUP2_CHLKE	Q39524 chlorella k
30	161	11.3	491	1 XYLE_ECOLI	P09098 escherichia
31	158.5	11.1	522	1 GTR2_RAT	P12336 rattus norv
32	158	11.1	534	1 HUP3_CHLKE	Q39525 chlorella k
33	157	11.0	435	1 YTI3_CAEEL	Q10917 caenorhabdi

34	156.5	11.0	464	1 GALP_ECOLI	P37021 escherichia
35	155	10.8	522	1 STA_RICCO	Q10710 ricinus com
36	154.5	10.8	584	1 ITRI_YEAST	P30605 saccharomyc
37	152	10.6	559	1 YAEC_SCHPO	Q09852 schizosacch
38	150.5	10.5	523	1 GTR2_MOUSE	P14246 mus musculu
39	149	10.4	612	1 ITR2_YEAST	P30606 saccharomyc
40	148.5	10.4	1222	1 YMP3_CAEEL	Q10947 caenorhabdi
41	148	10.4	472	1 EXUT_ECOLI	P42609 escherichia
42	146.5	10.3	546	1 HXT0_YEAST	P43581 saccharomyc
43	146.5	10.3	570	1 HXT1_YEAST	P32465 saccharomyc
44	145.5	10.2	422	1 EXUT_BACSU	O34456 bacillus su
45	145.5	10.2	524	1 GTR2_HUMAN	P11168 homo sapien

#### ALIGNMENTS

```

RESULT 1
ID Y0UL_CAEEL STANDARD; PRT; 529 AA.
AC P30638; Q21101;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III.
GN ZK637.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=92168156; PubMed=1538779;
RA Sulston J., Du Z., Thomas K., Wilson K., Hillier L., Staden R.,
RA Halloran N., Green P., Thierry-Mieg J., Oiu L., Dear S., Coulson A.,
RA Craxton M., Durbin R.K., Berns M., Metzstein M., Hawkins T.,
RA Alnscough K., Waterston K.;
RA "The C. elegans genome sequencing project: a beginning.";
RL Nature 356:37-41(1992).
[2]
RP REVISIONS.
RC STRAIN-BRISTOL N2;
RA Durbin R.;
RA Submitter (NOV-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z11115; CAA77460.1;
CC EMBL; Z22175; CAA77460.1; JOINED.
CC EMBL; Z22175; CAA80131.1;
CC EMBL; Z11115; CAA80131.1; JOINED.
CC PIR; S15786; S15786.
CC WormPep; ZK637.1; CE066638
CC InterPro; IPR003662; sub_trnsportr.
CC Pfam; PF00083; sugar_tr; 1.
CC PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
CC PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
CC KW Hypothetical protein; Transmembrane; Transport.
CC FT TRANSMEM 86 106 POTENTIAL.
CC FT TRANSMEM 122 142 POTENTIAL.
CC FT TRANSMEM 158 178 POTENTIAL.
CC FT TRANSMEM 238 258 POTENTIAL.
CC FT TRANSMEM 320 340 POTENTIAL.
CC FT TRANSMEM 373 393 POTENTIAL.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL; L05435; AAA42188.1; -

DR Pfam: PF00083; sugar\_tr; 1.  
KW Synapse; Nerve; Glycoprotein; Neurotransmitter transport;  
Transmembrane.  
FT DOMAIN 1 163 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 164 182 POTENTIAL.  
FT DOMAIN 183 204 LUMENAL (POTENTIAL).  
FT TRANSSEM 205 225 POTENTIAL.  
FT DOMAIN 226 242 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 243 260 POTENTIAL.  
FT DOMAIN 261 262 LUMENAL (POTENTIAL).  
FT TRANSSEM 263 281 POTENTIAL.  
FT DOMAIN 282 294 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 295 322 POTENTIAL.  
FT DOMAIN 323 334 LUMENAL (POTENTIAL).  
FT TRANSSEM 335 355 POTENTIAL.  
FT DOMAIN 356 445 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 446 469 POTENTIAL.  
FT DOMAIN 470 592 LUMENAL (POTENTIAL).  
FT TRANSSEM 593 611 POTENTIAL.  
FT DOMAIN 612 626 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 627 647 POTENTIAL.  
FT DOMAIN 648 649 LUMENAL (POTENTIAL).  
FT TRANSSEM 650 669 POTENTIAL.  
FT DOMAIN 670 694 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 695 711 POTENTIAL.  
FT DOMAIN 712 712 LUMENAL (POTENTIAL).  
FT TRANSSEM 713 731 POTENTIAL.  
FT DOMAIN 732 742 CYTOPLASMIC (POTENTIAL).  
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 742 AA; 82705 MW; 565DE7EF2929D5DB CRC64;

Query Match 23.9%; Score 341; DB 1; Length 742;  
Best Local Similarity 27.4%; Pred. No. 2,7e-18;  
Matches 86; Conservative 55; Mismatches 125; Indels 48; Gaps 5;

QY 1 EDVAEATGCGKFKQKLSVLTGLAWMADAMEMMILSIAPQLHCEWRLPSWOVALLTSVVF 60  
Db 153 ETLRECGHGFRQWTLTVFLGLALMADGVEVFGVFLPSAEKMDCLSDSNKMGMLIVY 212  
QY 61 VGMSSSTLMGNSIDQGRKTGKISVLTWLYGILSAFAPVSWILVRLGVLGFGIGG- 119  
Db 213 LGMVGAFLAGLADLRGRQCLLSISVNSVFAFFSFVQGYGTFLFCRLLSVGVIGGS 272  
QY 120 VPOSVTLYAEFLPMKARAKCILLIEVFWAIGTVFVVLAVFVMPISG----- 166  
Db 273 IPVFSYSEFLAQEKRGHLSWLCMFMMIGGVYAAAMAWAITPHYGWSFGMSAYOFHS 332  
QY 167 WRLLLSAVPLLLFVLCFLWPSARYDVLSGNOEKAIAIATLKRIATEN---GAP---M 219  
Db 333 WRVFLVFAFSPFAIGALTQPSRPFLENGKHDEAMVVLKQVHDTNRAKGHPERVF 392  
QY 220 PLGKLIIISRQEDR-----GKMRDLFTPHFRWTTLLWF 252  
Db 393 SVTHIKTHOEDELIEIQSDGTGTQWRGVRLSLGQGVWNGFLSCFSPRYRITLAMMG 452  
QY 253 IWFSNAFSYGLVL 266  
Db 453 VWFTMSFSYGLTV 466

RESULT 4  
PCAK\_PSEPU  
ID PCAK\_PSEPU STANDARD; PRT; 448 AA.  
AC Q51955;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE 4-HYDROXYBENZOATE TRANSPORTER.  
GN PCAK.  
OS Pseudomonas putida.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PRS2000;  
RX MEDLINE=95050205; PubMed=7961399;  
RA Harwood C.S., Nichols N.N., Kim M.-K., Ditty J.L., Parales R.E.;  
RT Identification of the psakKF gene cluster from Pseudomonas putida:  
involvement in chemotaxis, biodegradation, and transport of  
4-hydroxybenzoate".  
RL J. Bacteriol. 176:6479-6488(1994).  
CC -!- FUNCTION: TRANSPORTER FOR 4-HYDROXYBENZOATE. ALSO REQUIRED FOR  
CHEMOTAXIS TO AROMATIC ACIDS.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
(POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U10895; AAA85137.1; -  
DR InterPro: IPR003662; sub.trnsporttr.  
DR Pfam: PF00083; sugar\_tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Transport; Transmembrane; Inner membrane.  
FT TRANSSEM 31 51 POTENTIAL.  
FT TRANSSEM 51 51 POTENTIAL.  
FT TRANSSEM 58 88 POTENTIAL.  
FT TRANSSEM 95 115 POTENTIAL.  
FT TRANSSEM 120 140 POTENTIAL.  
FT TRANSSEM 153 173 POTENTIAL.  
FT TRANSSEM 185 205 POTENTIAL.  
FT TRANSSEM 262 282 POTENTIAL.  
FT TRANSSEM 302 322 POTENTIAL.  
FT TRANSSEM 330 350 POTENTIAL.  
FT TRANSSEM 352 372 POTENTIAL.  
FT TRANSSEM 399 419 POTENTIAL.  
FT TRANSSEM 422 442 POTENTIAL.  
SQ SEQUENCE 448 AA; 47176 MW; 6D5IC143123E99BC CRC64;

Query Match 19.8%; Score 282.5; DB 1; Length 448;  
Best Local Similarity 30.9%; Pred. No. 3.6e-14;  
Matches 83; Conservative 54; Mismatches 117; Indels 15; Gaps 8;

QY 11 KFWKLSVLTGLAWMADAMEMMILSIAPQLHCEWRLPSWOVALLTSVVFYGMSSSTLW 70  
Db 25 RYQWRVLLCFLVFLDGLDTAAMGFIAPALSQEWIDRASLGPVMSAALIGMVFGALGS 84  
QY 71 GNISDOYGRKTGLKISVLTWLYGILSAFAPVSWILVRLGVLGFGIG- GYPOSVTLYAE 129  
Db 85 GPLADRFGRKGVLVGAVLVFGGSLASAYATNDQLLVRLFTGLGLGAGNPATLLSE 144  
QY 130 FLPMKARAKCILLIEVF--WAIGTVFVFLAVFVMPISLGMWLLLSAV-PLLLFAVLVCF 186  
Db 145 YTP--ERLKSLLVTSMECGENLGNAGGGFISAKMIPAYGHSLLVIGVLPDLLALVLMV 202  
QY 187 WLPESARYDVL-SGNQEKAIAIATLKRIATENGAPMPLKGLIISRQE---DRCKMRDLFTPH 242  
Db 203 WLPESARFLVVRNRGTDKIRKTLSPAPQVVA--EAGSFSVPEQKAAARSFAVIESGT 260  
QY 243 FRWTTLLWFIFWFSNAFSYGLVLLTTEL 271  
Db 261 YGLGTMLLLTYFMGLVIVY---LLTSWL 286

RESULT 5

PCAK\_ACICA  
ID PCAK\_ACICA STANDARD; PRT; 457 AA.  
AC Q43975;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE 4-HYDROXYBENZOATE TRANSPORTER.  
GN PCAK.  
OS Acinetobacter calcoaceticus.  
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
OC Acinetobacter.  
OX NCBI\_TaxID=471;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BD413 / ADPL;  
RX MEDLINE=94341565; PubMed=8063101;  
RA Kowalchuk G.A., Harnett G.B., Benson A., Houghton J.E., Ngai K.-L.,  
RA Ornston L.N.;  
RT "Contrasting patterns of evolutionary divergence within the  
RT Acinetobacter calcoaceticus pca operon.";  
RL Gene 146:23-30(1994).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: L05770; AAC37151.1; -;  
CC InterPro: IPR003662; sub\_trnsportr.  
DR Pfam: PF00083; sugar\_tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Transport; Transmembrane; Inner membrane.  
FT TRANSMEM 35 55 POTENTIAL.  
FT TRANSMEM 73 93 POTENTIAL.  
FT TRANSMEM 102 122 POTENTIAL.  
FT TRANSMEM 129 149 POTENTIAL.  
FT TRANSMEM 169 189 POTENTIAL.  
FT TRANSMEM 192 212 POTENTIAL.  
FT TRANSMEM 275 295 POTENTIAL.  
FT TRANSMEM 311 331 POTENTIAL.  
FT TRANSMEM 339 359 POTENTIAL.  
FT TRANSMEM 365 385 POTENTIAL.  
FT TRANSMEM 401 421 POTENTIAL.  
FT TRANSMEM 427 447 POTENTIAL.  
SQ SEQUENCE 457 AA; 49277 MW; 4F5B5F77361A1567 CRC64;

Query Match 18.28; Score 260.5; DB 1; Length 457;  
Best Local Similarity 25.88; Pred. No. 1.6e-12;  
Matches 70; Conservative 64; Mismatches 120; Indels 17; Gaps 6;  
Qy 11 KFOKLSVLTGLAMMADAMEMMILSLAPQLHCEWRLPWSQVALLTSVVFVGMSSSTLW 70  
Db 32 RYOWLIAVCFLLIVFDGIDTAAMGFIAPALAOQDVGDRSOLGSPVMSAALGMIALVLS 91  
Qy 71 GNISDQYGRKTLKISVLTWLYGILSAFAPVYSWILVRLGLVGFIG-GVPSQVTLVAE 129  
Db 92 GPTADRFGRKIVLSMSLVFGFTLACAYSTNLDLSVIFRFLTGIGLGAAMPNATTLFSE 151  
Qy 130 FLPMKARAKCILLIEVFAIGTVEVVLAVFVMPSLGNRWLLILSA-VPLLLFVLCFWL 188  
Db 152 YCPAIRISLVTCEGYNLGMATGGFISSWLIFAFGWHSFLGLGGWAPLMLLVIFFL 211  
Qy 189 PESARYDVLG-NOEKATATLKRIA-----TENGAPMPLGKLIISQEDRGKMRDLFT 240  
Db 212 PESYRFLVKGKTKKVRQILSRAPQKVGQVTEFHPV---EEKVEAGTKKGFGVGLFS 267

Qy 241 PHFRWTTLLWFIMFSNAPSYGYGLVLLTTEL 271  
Db 268 AKYVKGTVLLWVTFMGLVMY---LLTSWL 295  
RESULT 6  
YVAJ\_BACSU  
ID YVAJ\_BACSU STANDARD; PRT; 451 AA.  
AC P37514;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YVAJ.  
GN YVAJ.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: D26185; BAA05214.1; -;  
CC EMBL: Z99124; CAB16121.1; -;  
CC Subtilist: BG10042; YVAJ.  
DR InterPro: IPR003662; sub\_trnsportr.  
DR Pfam: PF00083; sugar\_tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; FALSE.NEG.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; FALSE.NEG.  
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.  
FT TRANSMEM 30 50 POTENTIAL.  
FT TRANSMEM 61 81 POTENTIAL.  
FT TRANSMEM 99 119 POTENTIAL.  
FT TRANSMEM 128 148 POTENTIAL.  
FT TRANSMEM 159 179 POTENTIAL.  
FT TRANSMEM 183 203 POTENTIAL.  
FT TRANSMEM 271 291 POTENTIAL.  
FT TRANSMEM 306 326 POTENTIAL.  
FT TRANSMEM 345 365 POTENTIAL.  
FT TRANSMEM 397 417 POTENTIAL.  
FT TRANSMEM 422 442 POTENTIAL.  
SQ SEQUENCE 451 AA; 49672 MW; 4A00D3891C4D7D09 CRC64;

Query Match 17.08; Score 242.5; DB 1; Length 451;  
Best Local Similarity 27.48; Pred. No. 3.3e-11;  
Matches 75; Conservative 49; Mismatches 129; Indels 21; Gaps 7;  
Qy 4 VEATGFGFQWKLSTGLAMMADAMEMMILSLAPQLHCEWRLPWSQVALLTSVVFVGM 63  
Db 20 LDRPISRVHQVLTALGIVYFFDLADLFTLSNVAPALIEHWGIPLSLIANYTAASFLGM 79  
Qy 64 MSSSTLWGNISDQYGRKTLKISVLTWLYGILSAFAPVYSWILVRLGLVGFIGVPS 123  
Db 80 FLGASLGLGRSLSDRIGRKKALNLFVVFVSTASLCNAAMDIPSLMTFRFLTGFGVAAAWY 139  
Qy 124 VTLY-AEFLPMKARAKCILLIEVFAIGTVEVVLAVFVMP--SLGNRWLLILSAVPLLL 180









```

Db 253 LFAEILCRPLLGTVLLMGVFF 274
::
|||||::|
RESULT 11
ID OCN2_MOUSE STANDARD; PRT; 557 AA.
AC Q920E8;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTTRANSPORTER).
GN SLC22A5 OR OCTN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=99113835; PubMed=9916797;
RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,
RA Nikaido H., Sai Y., Kozumi A., Shoji Y., Takada G., Matsuishi T.,
RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,
RA Tsuji A.;
RT "Primary systemic carnitine deficiency is caused by mutations in a
RT gene encoding sodium ion-dependent carnitine transporter.";
RL Nat. Genet. 21:91-94(1999).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANT JVS ARG-352.
RC STRAIN=C3H;
RX MEDLINE=99057546; PubMed=9837751;
RA Lu K., Nishimori H., Nakamura Y., Shima K., Kuwajima M.;
RT "A missense mutation of mouse OCTN2, a sodium-dependent carnitine
RT cotransporter, in the juvenile visceral steatosis mouse.";
RL Biochem. Biophys. Res. Commun. 252:590-594(1998).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99384224; PubMed=10454528;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
RA Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic
RT cation transporter 2 (OCTN2), an organic cation/carnitine
RT transporter.";
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
CC -!- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CC CARNITINE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF JUVENILE VISCERAL
CC STEATOSIS (JVS).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CC CATION SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB015800; BAA36590.1; -
DR EMBL; AF111425; AAC99787.1; -
DR EMBL; AF110417; AAD54060.1; -
DR MGD; MGI:1329012; SLC22a5.
DR InterPro; IPR003662; sub_transprot.
DR Pfam; PF00083; sugar_tr; 1.
DR ProSITE; PS00216; SUGAR_TRANSPORT_1; 1.
KW Transport; Transmembrane; Glycoprotein; Disease mutation.
FT TRANSMEM 21
POTENTIAL.

```

```

FT TRANSMEM 143 163
FT TRANSMEM 173 193
FT TRANSMEM 198 218
FT TRANSMEM 233 253
FT TRANSMEM 258 278
FT TRANSMEM 342 362
FT TRANSMEM 374 394
FT TRANSMEM 407 427
FT TRANSMEM 431 451
FT TRANSMEM 489 509
FT CARBOHYD 57 57
FT CARBOHYD 64 64
FT CARBOHYD 91 91
FT CARBOHYD 322 322
FT VARIANT 352
FT SEQUENCE 557 AA; 62779 MW; 6093F0EE9612B204 CRC64;
SQ
Query Match 15.3%; Score 219; DB 1; Length 557;
Best Local Similarity 27.9%; Pred. No. 2.2e-09;
Matches 69; Conservative 43; Mismatches 107; Indels 28; Gaps 6;
QY 44 EHRL-----PSQVALLTSVVFVGMSSSTLWGNISDOYGRKTGLKISVLWTLTYGILSAPA 100
Db 131 EMDLVCKDDMKAPLTLSLFFVGLMGFSISGQLSDRFGKRNVLFLTMGMQTFGSFLOVFS 190
QY 101 PYVSWILVLRGLVGFG-IGVQPQSVTLYAEFLPMKARAKCILL-IEVFWAIGTVFEVFLA 158
Db 191 VNFEMFTVLFLVGMGQISNYVAAFVLGTEILSKSRIIFATLGCVLGFYAFGFWPLPFA 250
QY 159 VFVPSLGRWRLILSAVPLLLFAVLGFWLPESARYDVLSGNOEKAIAIUKRIATENGAP 218
Db 251 YFIR---DWRMLLLALTVPGLGALWGFIPESPRWLISQGRKEAEVIRKAAKINGIV 307
QY 219 MPLGLKLIISROEDRGKMRDL-----FPHFRWTTLLWFWFSNAFSYGL 264
Db 308 AP-----STIFDPSELQDLNLTSTKPOLHHYDILIRNIRVITMSILMLTISVGFGFL 361
QY 265 VLLTTEL 271
Db 362 SLDTPLN 368

```

```

RESULT 12
ID OCN2_MOUSE STANDARD; PRT; 557 AA.
AC Q70594; Q9QWLO;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTTRANSPORTER)
DE (JVS2R) (CT1).
GN SLC22A5 OR OCTN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney.
RX MEDLINE=98200080; PubMed=9541011;
RA Schoemig E., Spitzenberger F., Engelhardt M., Martel F., Oerding N.,
RA Gruendemann D.;
RT "Molecular cloning and characterization of two novel transport
RT proteins from rat kidney.";
RL FEBS Lett. 425:79-86(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;
RX MEDLINE=99011422; PubMed=9792817;
RA Sekine T., Kusuhabara H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y.,
RA Kanai Y., Endou H.;

```



QY 4 VEAIGFGKFWKLSVLTGLAWMADAMMMILSILAPQLHCEWRLPSWQVALLTSVVFVGM 63  
Db 8 MDDLPLNRHCHIAALTFCAGHTDGYVLGVIGYAIQLTPAMQLTPFMAGMIGGSSALLGL 67  
QY 64 MSSTLWGNISDYGKRGKLGKLSVLTWLYGILSAFAPYVSWILVRLGLVGFVGIGVQPS 123  
Db 68 FLGSLVLGWSIDHIGRQKFTFSFLITLILASLFOFFATTPEHLILIGLIGLGG-DYS 126  
QY 124 V--TLVAFELPMKARAKCILLIEVFAIGTVEVVLAVFVMPSL-----GNRWL 170  
Db 127 VGHITLAEFSPRRHGIILGAFSVWTVG-----YVLASIAGHFISENPANRWL 177  
QY 171 LILSAVPLLLFVAFVLPESARYDVLNQEKATATKRIATENGAPMLGKLIIISROE 230  
Db 178 LASAALPALLITLIRWGTPTSPRWLLRQGRFAEHAIVHRYE---GPHVLLGDVVVTAH 234  
QY 231 DRCK-----MRDLTFPHFWTLLWFI---WFSNAFSYYGL 264  
Db 235 KHIKTLFSSRYRRRTAFNSVFFVCLVIPWFVIYTWLPTIAQTIGL 279

RESULT 14  
MMH\_ALCEU STANDARD; PRT; 428 AA.  
AC 051798;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PROBABLE 4-METHYLMUCONOLACTONE TRANSPORTER.  
GN MMH.  
OS Alcaligenes eutrophus (Ralstonia eutrophus).  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI\_TaxID=510;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JMP134;  
RX MEDLINE=98121313; PubMed=9461415;  
RA ERB R.W., Timmis K.N., Pieper D.H.;  
RT "Characterization of a gene cluster from Ralstonia eutrophia JMP134 encoding metabolism of 4-methylmucunolactone.";  
RL Gene 206:53-62(1998).  
CC -!- FUNCTION: PROBABLE UPTAKE OF 4-METHYLMUCONOLACTONE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X99639; CAA67957.1; -  
DR InterPro: IPR003662; sub.trnsport.  
DR Pfam: PF000083; sugar.tr.1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Transport; Transmembrane.  
FT TRANSMEM 27 47 POTENTIAL.  
FT TRANSMEM 51 71 POTENTIAL.  
FT TRANSMEM 81 101 POTENTIAL.  
FT TRANSMEM 111 131 POTENTIAL.  
FT TRANSMEM 146 166 POTENTIAL.  
FT TRANSMEM 168 188 POTENTIAL.  
FT TRANSMEM 228 248 POTENTIAL.  
FT TRANSMEM 253 273 POTENTIAL.  
FT TRANSMEM 288 308 POTENTIAL.  
FT TRANSMEM 315 335 POTENTIAL.  
FT TRANSMEM 357 377 POTENTIAL.  
FT TRANSMEM 378 398 POTENTIAL.  
SQ SEQUENCE 428 AA; 45804 MW; 7CC9096F6F7BE230 CRC64;

Query Match 14.2%; Score 203.5; DB 1; Length 428;  
Best Local Similarity 27.4%; Pred. No. 2.5e-08;  
Matches 49; Conservative 38; Mismatches 83; Indels 9; Gaps 4;  
QY 24 WNADEMWMILSILAPQLHCEWRLPSWQVALLTSVVFVGMSSSTLWGN--ISDQYGRKT 81  
Db 23 WALDSFDMQMFSLPALTLTWGLTKAEVGLGTVALV--VTAIGWAGAGILSDRYGRAR 80  
QY 82 CLKISVLTWLYGILSAFAPYVSWILVRLGLVGFVGIGV-POSVTLYAEFLPMKARAKCI 140  
Db 81 ILVLAIITWTLFGVLAGFAQSTQQQLIARTLOGLFGGGEWAYCAALMAEVIDSRHKGAI 140  
QY 141 LLIE-----VFVAIGTVFEVVLAVFVMPSLGWRMLLILSAVPLLLFVAFVLPESARYD 195  
Db 141 GFVQSGFALGVALVAVVATLLLAWLPKEMARVAFWSGIIPALIVLFIIRHVKSSMFE 199  
RESULT 15  
STP1\_ARATH STANDARD; PRT; 522 AA.  
AC P23586;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE GLUCOSE TRANSPORTER (SUGAR CARRIER).  
GN STP1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. LANDSBERG ERRECTA;  
RX MEDLINE=91005995; PubMed=2209537;  
RA Sauer N., Friedlaender K., Graeme-Wicke U.;  
RT "Primary structure, genomic organization and heterologous expression of a glucose transporter from Arabidopsis thaliana.";  
RL EMBO J. 9:3045-3050(1990).  
CC -!- FUNCTION: ACTIVE UPTAKE OF HEXOSES. PROBABLE GLUCOSE/HYDROGEN SYMPORT.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X55350; CAA39037.1; -  
DR PIR; S12042; S12042.  
DR InterPro: IPR003662; Sugar\_trnsport.  
DR InterPro: IPR003662; sub.trnsport.  
DR Pfam: PF000083; sugar.tr.1.  
DR PRINTS; PR00171; SUGTRNSPORT.  
DR PRINTS; PR00172; GLUCTRNSPORT.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Transmembrane; Transport; Sugar transport; Symport.  
FT DOMAIN 1 21 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 22 42 POTENTIAL.  
FT TRANSMEM 84 104 POTENTIAL.  
FT TRANSMEM 113 133 POTENTIAL.  
FT TRANSMEM 141 161 POTENTIAL.  
FT TRANSMEM 168 188 POTENTIAL.  
FT TRANSMEM 200 220 POTENTIAL.  
FT TRANSMEM 285 305 POTENTIAL.  
FT TRANSMEM 321 341 POTENTIAL.

```

FT TRANSMEM 349 369 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
FT DOMAIN 474 522 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 522 AA; 57596 MW; 68A6C72AFFD90380 CRC64;

Query Match 14.0%; Score 200; DB 1; Length 522;
Best Local Similarity 24.0%; Pred. No. 5.3e-08;
Matches 69; Conservative 59; Mismatches 135; Indels 24; Gaps 9;

Qy 4 VEAIGFGKFWKLSVLTGLAWMADAMEMMILSILAPQ-----LHCEWRLPNQVALLT 56
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 29 VAAMGGLIFGYDIGSGVTSMPSEFLARFFPSVYRKQOEDASTNQYCOYDSPT--LTMFT 86
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 57 SVFVGMSSSTLWGNISDQYGRKTGLKISVLWTLYCYILSAPAPYYSWILVL-RGLVGF 115
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 87 SSLYLAALISLVASTVTRKFGRLSMLFGGILFCAGALINGFAK-HVWMLIVGRILLGF 145
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 116 GIGGVPSQVTLY-AEFLPMKARAKCILLIEVFWAIG-TVEVVLAVFVMPSLGWRLLIL 173
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 146 GIGFANQAVPLYLSEMAPYKYRGALNIGFQLSITIGILVAEVLNFFAKIKGGWRLSL 205
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 174 --SAPVLLLFVLCFWLPESARYDVLSGNOEKAATLKRATENGAPMPGLKLIISQED 231
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 206 GGAVVPALIIITIGSLVLPDTPNSMIERQHEEAKTKLRRIRGVDDVDSOEEDDLVAASKES 265
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 232 -----RGKWRDLFTPHFRWTTLLLWFLWFS--NAFSYYGLVLLTT 269
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 266 QSIHPWRNLLRRKRYRPHLTMAVMIPFFQQLTGINVMFYAPVLENT 312
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: March 13, 2002, 12:42:53  
Job time: 159 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:42:29 ; Search time 43.34 Seconds  
(without alignments)  
921.374 Million cell updates/sec

Title: US-09-911-667A-4  
Perfect score: 1429  
Sequence: 1 EDAVEAIGFGKFWKLSVLT.....WFSNAPSYYGLVLLTTELPQ 273

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL17:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1409	98.6	548	11 Q92217	Q92217 rattus norv
2	751.5	52.6	497	5 Q9W1D4	Q9W1D4 drosophila
3	617.5	43.2	470	10 Q9LSH7	Q9LSH7 arabidopsis
4	609.5	42.7	500	10 Q9M7W2	Q9M7W2 arabidopsis
5	446.5	31.2	454	2 Q9RYN9	Q9RYN9 deinococcus
6	358.5	25.1	422	2 Q9X1T9	Q9X1T9 thermotoga
7	344	24.1	727	11 Q9Z216	Q9Z216 rattus norv
8	342	23.9	742	11 Q9J1S5	Q9J1S5 mus musculu
9	341.5	23.9	709	5 Q9W3W9	Q9W3W9 drosophila
10	341	23.9	724	13 Q9O4O6	Q9O4O6 discopyge o
11	338	23.7	607	4 Q9BVZ9	Q9BVZ9 homo sapien
12	338	23.7	742	4 Q9A841	Q9A841 homo sapien
13	338	23.7	742	6 Q9397	Q9397 bos taurus
14	331	23.2	683	4 Q9A840	Q9A840 homo sapien
15	320	22.4	455	2 Q9HYD9	Q9HYD9 pseudomonas
16	320	22.4	683	11 Q63564	Q63564 rattus norv
17	298	20.9	443	2 Q9L7Y1	Q9L7Y1 pseudomonas
18	281	19.7	556	11 Q63089	Q63089 rattus norv
19	281	19.7	593	11 P70485	P70485 rattus norv

20	280	19.6	555	11 Q9ROW2	Q9ROW2 rattus norv
21	280	19.6	593	11 P97558	P97558 rattus norv
22	278	19.5	536	11 Q9RIU7	Q9RIU7 rattus norv
23	277	19.4	553	11 Q70577	Q70577 mus musculu
24	275.5	19.3	556	11 Q08966	Q08966 mus musculu
25	275	19.2	554	6 Q02713	Q02713 sus scrofa
26	275	19.2	556	11 Q9RIQ4	Q9RIQ4 mus musculu
27	273	19.1	555	4 Q15244	Q15244 homo sapien
28	271.5	19.0	535	11 Q63314	Q63314 rattus norv
29	270.5	18.9	537	11 Q88909	Q88909 mus musculu
30	267.5	18.7	448	2 Q9I6Q3	Q9I6Q3 pseudomonas
31	267.5	18.7	562	13 Q57379	Q57379 pseudopleur
32	265	18.5	446	2 Q9HUR3	Q9HUR3 pseudomonas
33	264	18.5	448	2 Q9I110	Q9I110 pseudomonas
34	264	18.5	526	10 Q9LHQ6	Q9LHQ6 arabidopsis
35	258	18.1	465	2 Q9AQ12	Q9AQ12 pseudomonas
36	257	18.0	548	5 Q01384	Q01384 drosophila
37	257	18.0	548	5 Q9VCA2	Q9VCA2 drosophila
38	257	18.0	554	4 Q15395	Q15395 homo sapien
39	255	17.8	554	4 Q15245	Q15245 homo sapien
40	251	17.6	554	6 Q77504	Q77504 oryctolagus
41	249	17.4	551	4 Q9Y226	Q9Y226 homo sapien
42	248.5	17.4	463	2 Q9RHQ1	Q9RHQ1 variovorax
43	247	17.3	554	4 Q9NQD4	Q9NQD4 homo sapien
44	246	17.2	567	5 Q9VCA3	Q9VCA3 drosophila
45	245	17.1	430	11 Q35882	Q35882 rattus norv

ALIGNMENTS

RESULT 1  
Q92217  
ID Q92217 PRELIMINARY; PRT; 548 AA.  
AC Q92217;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE SV2 RELATED PROTEIN.  
GN SVOP.  
OS Rattus norvegicus (rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX MEDLINE=99019745; PubMed=9801366;  
RA Janz R., Hofmann K., Sudhof T.C.;  
RT "SVOP, an evolutionarily conserved synaptic vesicle protein, suggests novel transport functions of synaptic vesicles.";  
RL J. Neurosci. 18:9269-9281(1998).  
CC -!- SURCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL; AF060173; AAC78627.1; -;  
DR InterPro; IPR003662; sub\_transporter.  
DR Pfam; PF00083; sugaf\_tr; 1.  
KW Transmembrane.  
SQ SEQUENCE 548 AA; 60804 MW; 19AD8475B7579496 CRC64;

Query Match 98.6%; Score 1409; DB 11; Length 548;  
Best Local Similarity 97.4%; Pred. No. 1.3e-91;  
Matches 266; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
QY 1 EDAVEAIGFGKFWKLSVLTGLAWMADAMEMILSLAPOLHCEWRLPSQVALLTSVVF 60  
DB 71 EDAVEAIGFGKFWKLSVLTGLAWMADAMEMILSLAPOLHCEWRLPSQVALLTSVVF 130  
QY 61 VGMSSSTLWGNISDQYGRKTKGLKISVLWTLYGILSAFAPYVSWTLVLRLGVGFGIGV 120  
DB 131 IGMSSSTLWGNISDQYGRKTKGLKISVFTLYYGLSAFAPYVSWTLVLRLGVGFGIGV 190



QY 121 POSVTLYAEFLPMKARAKACILLIEVFMAIGTVFVAVFVMPSLGWRWLLILSAVPLLL 180  
DB 191 POSVTLYAEFLPMKARAKACILLIEVFMAIGTVFVAVFVMPSLGWRWLLILSAVPLLL 250  
QY 181 FAVLCFWLPESARYDVLGNSQEKATATLKRIATENGAPMLGKLIISROEDRGKMRDLFT 240  
DB 251 FAVLCFWLPESARYDVLGNSQEKATATLKRIATENGAPMLGKLIISROEDRGKMRDLFT 310  
QY 241 PFRWTTLLWFIWFSNAFSYYGLVLLTTEL 273  
DB 311 PFRWTTLLWFIWFSNAFSYYGLVLLTTEL 343

RESULT 2  
Q9W1D4 PRELIMINARY; PRT; 497 AA.  
AC Q9W1D4;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE CG4324 PROTEIN.  
GN CG4324.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Phryganea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle E.G., Heit E.G., Nelson C.R., Miklos G.B.G.,  
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Balles R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berlan B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iyegawa C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Nusskern D.R., Pacleb J.M.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RT Science 287:2185-2195(2000).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL; AE003462; AAF47135.1; -;  
DR FlyBase; FBgn0034956; CG4324.

DR InterPro: IPR003662; sub\_transporter.  
DR Pfam; PF00083; sugar\_tf; 1.  
KW Transmembrane.  
SQ SEQUENCE 497 AA; 55360 MW; 678F44A3286B7B04 CRC64;

Query Match 52.6%; Score 751.5; DB 5; Length 497;  
Best Local Similarity 53.7%; Pred. No. 2e-45;  
Matches 146; Conservative 46; Mismatches 79; Indels 1; Gaps 1;

QY 1 EDVAEIAFGKFWKLVLTGLAWMADAMEMMILSILAPOLHCEWRLPSQVALLTSVVF 60  
DB 62 QQAIAAFGFWFHVKLSLLVGLGWNDSMEMAILSLGSLFCEWNVTKFQQAASVTTVVF 121  
QY 61 VGMSSSTLWGNISDQYGRKTKGLKISVLTWLVYVILSFAFVYSWILVLGLVGFICGV 120  
DB 122 LGHMLSSSTFWTQLSNRYGRKKSALTIFGVLLVLYLSLVSVAPSYAWLLTLRGLVGAIGCV 181  
QY 121 POSVTLYAEFLPMKARAKACILLIEVFMAIGTVFVAVFVMPSLGWRWLLILSAVPLLL 180  
DB 182 POSVTLYAEFLPTKHGKCVLMDCFWALGACFEVVLAVLVYPYVGWRLLSALSATPLLI 241  
QY 181 FAVLCFWLPESARYDVLGNSQEKATATLKRIATENGAPMLGKLIISROEDRGK-MRDLF 239  
DB 242 FTILSPWSESARYSYNGHNDKAIKVLQIAHNKRHMLMGLMADDEPSCAESFRSL 301  
QY 240 TPFHRTWTTLLWFIWFSNAFSYYGLVLLTTEL 271  
DB 302 SPFLYRTTLLWFLWLASAFYVGLVLTTEL 333

RESULT 3  
Q9LSH7 PRELIMINARY; PRT; 470 AA.  
AC Q9LSH7;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE TRANSPORTER-LIKE PROTEIN.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=20277480; PubMed=10819329;  
RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence features of the regions of 4,504,864 bp covered by sixty pl and TAC clones.";  
RL DNA Res. 7:131-135(2000).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL; AB026645; BAB02515.1; -;  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam; PF00083; sugar\_tf; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 470 AA; 51457 MW; 1062413BA354616D CRC64;

Query Match 43.2%; Score 617.5; DB 10; Length 470;  
Best Local Similarity 43.5%; Pred. No. 5e-36;  
Matches 121; Conservative 52; Mismatches 98; Indels 7; Gaps 1;

QY 1 EDVAEIAFGKFWKLVLTGLAWMADAMEMMILSILAPOLHCEWRLPSQVALLTSVVF 60

Db 11 DEALVAMGCKFQIYVLAYAGMGWAEAMEMMLLSFVGPVAVQSLWNLSARQESLITSVF 70  
Qy 61 VGMSSSTLWGNISDOYGRKTKLISVLTYGILSAFAPYVSWILVRLGLVGFGGV 120  
Db 71 AGMLIGAYSWGIVSDKHRRKGFIIITAVVTFVAGFSAFSPNYMMLIILCLVGLGCGG 130  
Qy 121 POSVTLAAEFLPMKARAKCILLIEVFVMAIGTVFEVVLAVFVMPSLGWRWLLILSAVPLLL 180  
Db 131 PVLASWYLEFIPAPSRCTWVWVFAFWTGTIFEASLAWLMPRLGWRWLLAFSSVPSL 190  
Qy 181 FAVLCFWLPESARYDVLSGNOEKATLKRATENGAPMPLGK-----IISRQEDRG 233  
Db 191 LLLFYRWTSFPRYLILQGRKAELALEKRIARNKTKQLPGLVSSSELETELENKPGF 250  
Qy 234 KMRDLFTPHFRWTTLLWFWFNSAFSYGVLVLLTTTEL 271  
Db 251 SLLALLSPTLMKRTLLWVFFGNFAYYGVVLLTTTEL 288

RESULT 4  
Q9M7W2 ID Q9M7W2 PRELIMINARY; PRT; 500 AA.  
AC Q9M7W2; 2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE PUTATIVE TRANSPORTER.  
GN MG6.16.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,  
RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,  
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
RA "Arabidopsis thaliana chromosome III P1 MGH6 genomic sequence";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RC sub-CELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL; AC024128; AAF35954.1; -.  
DR InterPro; IPR003662; sub.transporter.  
DR Pfam; PF00083; sugar\_tr; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
DR Transmembrane.  
KW Transmembrane.  
SQ SEQUENCE 500 AA; 54573 MW; EC459EB09EF581C5 CRC64;

Query Match 42.7%; Score 609.5; DB 10; Length 500;  
Best Local Similarity 39.3%; Pred. No. 2e-35;  
Matches 121; Conservative 55; Mismatches 95; Indels 37; Gaps 2;  
Qy 1 EDVAIGFGKFWKLSVLTGLAMMADAMEMMLISILAPQLHCEWRLPSQVALLTSVVF 60  
Db 11 DEALVAMGFGFQIYVLAYAGMGWAEAMEMMLLSFVGPVAVQSLWNLSARQESLITSVF 70  
Qy 61 VGMSSSTLWGNISDOYGRKTKLISVLTYGILSAFAPYVSWILVRLGLVGFGGV 120  
Db 71 AGMLIGAYSWGIVSDKHRRKGFIIITAVVTFVAGFSAFSPNYMMLIILCLVGLGCGG 130  
Qy 121 POSVTLAAEFLPMKARAKCILLIEVFVMAIGTVFEVVLAVFVMPSLGWRWLLILSAVPLLL 180  
Db 131 PVLASWYLEFIPAPSRCTWVWVFAFWTGTIFEASLAWLMPRLGWRWLLAFSSVPSL 190  
Qy 181 FAVLCFWLPESARYDVLSGNOEKATLKRATENGAPMPLG-----KLIISRQEDRG 222  
Db 191 LLLFYRWTSFPRYLILQGRKAELALEKRIARNKTKQLPGLVSSSELETELENKNIPT 250  
Qy 223 -----KLIISRQEDRG-KMRDLFTPHFRWTTLLWFWFNSAFSYG 263

Db 251 ENTHLLKAGSGEAVAVSKIVLAKRDEPGFSLLALLSPTLMKRTLLWVFFGNFAYY 310  
Qy 264 LVLLTTTEL 271  
Db 311 VVLLTTTEL 318  
RESULT 5  
Q9RYN9 ID Q9RYN9 PRELIMINARY; PRT; 454 AA.  
AC Q9RYN9;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE SUGAR TRANSPORTER, PUTATIVE.  
GN DRA0271.  
OS Deinococcus radiodurans.  
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
OX NCBI\_TaxID=1299;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=R1;  
RX MEDLINE=20036896; PubMed=10567266;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamlile W., Crosby M., Shen M.,  
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C., D.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans R1";  
RL Science 286:1571-1577(1999).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL; AE001863; AAF12486.1; -.  
DR TIGR; DRA0271; -.  
DR InterPro; IPR003662; sub.transporter.  
DR Pfam; PF00083; sugar\_tr; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
KW Complete proteome; Sugar transport; Transmembrane.  
SQ SEQUENCE 454 AA; 48171 MW; 214EA1A3EDC60B8B CRC64;

Query Match 31.2%; Score 446.5; DB 2; Length 454;  
Best Local Similarity 35.1%; Pred. No. 5.1e-24;  
Matches 94; Conservative 52; Mismatches 115; Indels 7; Gaps 3;  
Qy 4 VEAIGFGKFWKLSVLTGLAMMADAMEMMLISILAPQLHCEWRLPSQVAVV 61  
Db 19 LDDGLGRFQWKLKLLAICGLTWAADAMEVLLMGFALPGISAAPFKGSPAATMLLTATFA 78  
Qy 62 GMMSSSTLWGNISDOYGRKTKLISVLTYGILSAFAPYVSWILVRLGLVGFGGV-V 120  
Db 79 GMLFGMFWGLADRVGRSRVFLTVAGLVVFGAGALAPTLLTWLLVAFRLTGAIGGTL 138  
Qy 121 POSVTLAAEFLPMKARAKCILLIEVFVMAIGTVFEVVLAVFV---MPSLGRWLLILSAV 176  
Db 139 PVDYSMAAEFVPTAWRGFVLVLESFNAVGIVVVAALAWWSTAFAPAEGRWLLGLAAL 198  
Qy 177 PLLLFVLCFWLPESARYDVLSGNOEKATLKRATENGAPMPLGKLIISRQEDRGMR 236  
Db 199 PGLVGLIARIIGIPDSRSLARGEAEQARAALQKVAQANGGTLPAPLAHPQEPQPRVSPA 258  
Qy 237 DLFTPHFRWTTLLWFWFNSAFSYG 264  
Db 259 QLFRGVLAARRPTLLWVTFGLSLGYGI 286  
RESULT 6  
Q9X1T9 ID Q9X1T9 PRELIMINARY; PRT; 422 AA.

```
AC O9X1T9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PERMEASE, PUTATIVE.
GN TM1603.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN RP
SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AE001804; AAD36670.1; -.
DR TIGR; TM1603; -.
DR InterPro; IPR00515; BPD_transp.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr_1.
DR PROSITE; PS00402; BPD_TRANS_INN_MEMBER; UNKNOWN_1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Complete proteome; Transmembrane.
SQ SEQUENCE 422 AA; 47152 MW; 13f9DC5649A1338D CRC64;

Query Match 25.1%; Score 358.5; DB 2; Length 422;
Best Local Similarity 32.08; Pred. No. 7.4e-18;
Matches 89; Conservative 55; Mismatches 111; Indels 23; Gaps 7;

QY 1 EDVAEIGFGKQKLSVLTGLWADAMENMILSILAPQLHCEWRLPSQVALLTSVVF 60
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 4 DEIVEKYVDKRTQRFILTSIAWFDAAAGVLLSFVLPIYKWNLTSTQCATIASATF 63

QY 61 VGMSSSTLWGNISDQYGRKTKLSVLWTLTYGILSAFAPYYSWILVRLGLVFGIGV 120
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 64 LGMLFGALSGVEVDLLGRKVSNNLFFIVITITFLSGFSSFFETLLRLGLSGFGYGL 123

QY 121 PQSVTLY-AEFLPMKARAKCILLIEVFWAIGTVFEWLVAVFVMPSLGWRLLILSAVPLL 179
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 124 MFSFNAYLAFTSIRLGRYLVLLESSNAVGSILIGLFAVNVLPN-WRWVFWIFSIG-Y 180

QY 180 LFAVLGFWLPESARYDVLSGNOKAIAIYL-KRIATENGAP-----MPLGKLIISROEDRGK 234
   | | :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 181 LFPVFLRMPETPKYAFLLGKGKALESLGRKVEEVELPKKPKVPILALL-----KREH 235

QY 235 MRDLTPHPRWTTLLLWFIWISNAPSYGYGLVLLTTELF 272
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 236 LKD-----TVVILWIAFWVSVVYALFTWAPRIF 264

RESULT 7
Q9Z2I6 PRELIMINARY; PRT; 727 AA.
AC Q9Z2I6;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE SYNAPTIC VESICLE PROTEIN 2C.
GN SV2C.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
```

```
RN RP
SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99019745; PubMed=9801366;
RA Janz R., Hofmann K., Sudhof T.C.;
RT "SVOP, an evolutionarily conserved synaptic vesicle protein, suggests
RT novel transport functions of synaptic vesicles.";
RL J. Neurosci. 18:9269-9281(1998).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AF060174; AAC78628.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr_1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 727 AA; 82248 MW; 4C154C69341D8DB7 CRC64;

Query Match 24.1%; Score 344; DB 11; Length 727;
Best Local Similarity 28.4%; Pred. No. 1.3e-16;
Matches 88; Conservative 57; Mismatches 115; Indels 50; Gaps 7;

QY 4 VEAIGFGKQKLSVLTGLWADAMENMILSILAPQLHCEWRLPSQVALLTSVVFVGM 63
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 142 IOECGCHGRFPQWALFFVLGMLMADGVEFVGVFLPSAETDLCIPNSGSLGSIYVLGM 201

QY 64 MSSSTLWGNISDQYGRKTKLSVLWTLTYGILSAFAPYYSWILVRLGLVFGIGG-VPO 122
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 202 MVGAFFWGGGLADKVGKQSLICMSVNGVFPAFLSSVQGYGFFLLCRLLSGFGIGAIPT 261

QY 123 SVTLYAEFLPMKARAKCILLIEVFWAIGTVFEWLVAVFVMPSLG-----WRW 169
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 262 VFSYFAEVLAREKRGHLSWLCMFWMIGGIYASAMAWAIIPIHYGWSFGMSYAFQHSWRV 321

QY 170 LLILSAVPLLLFAVLGFWLPESARYDVLSGNOKAIAIYL-KRIATEN-----MPLG 222
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 322 FVIVCALPCVSVVALTFMPESPRLLELVGKHDEAWMLKLHDTNMRARGQPEKVFVN 381

QY 223 KL-----ILSRQEDRGK-----MRDLTPHPRWTTLLLWFIW 254
   | :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 382 KIKTPQKIDELIEISDTGTWYRRCFVRIRTELYGIWLTFR-CFNPVPRENTIKLTIVW 440

QY 255 FSNAPSYGYGL 264
   | :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 441 FTLSFGYYGL 450

RESULT 8
Q9JIS5 PRELIMINARY; PRT; 742 AA.
AC Q9JIS5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CA2+ REGULATOR SV2A.
GN SV2A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN RP
SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=20088300; PubMed=10624962;
RA Janz R., Goda Y., Geppert M., Missler M., Sudhof T.C.;
RT "SV2A and SV2B function as redundant Ca2+ regulators in
RT neurotransmitter release.";
RL Neuron 24:1003-1016(1999).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AF196781; AAF87321.1; -.
DR EMBL; AF196780; AAF87321.1; JOINED.
DR MGI; MGI:1927139; SV2a.
```

```

DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR PROSITE: PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 742 AA: 82647 MW: 1074857FD13ED894 CRC64;

Query Match 23.9%; Score 342; DB 11; Length 742;
Best Local Similarity 27.4%; Pred. No. 1.9e-16;
Matches 86; Conservative 55; Mismatches 125; Indels 48; Gaps 5;

Qy 1 EDVAEAIQFGKFWKLSVLTGLAWMADAMEMMILSLAPQLHCEWRLPSMQVALLTSVVF 60
Db 153 ETILRECHGHEFWTLFVGLGALMADGVEFVVGFLPSAEKDWCLSDSKNGMLGLIVY 212
Qy 61 VGMSSSTLWGNISDQYGRKTKLISVLTLYYGLSFAFVYVSWILVRLGVFGIGV 119
Db 213 LGMVGFALWGLADRLGRQCLLSISVNSVFAFFSFVGYCTFLFCRLLSGVGIGGS 272
Qy 120 VPOSVTLAEFLPMKARAKCILLTEVFAIGTVFEVVLAVFVMPSLG----- 166
Db 273 IPVFSYSEFLAEQKKEHLSWLCFMWVGCVYAAAMAWAIIHYGWSFOMK;SAYQFHS 332
Qy 167 WRWLILSAVPLLLFAVLCFELPESARYDYLSGNOEKAIATLKRIATEN-----M 219
Db 333 WRVFLVCAFFSVAIGALTTPESPFRFFLENGKHDEAWMLKVQVHDTNMRKAGHPERV 392
Qy 220 PLGKLIISROEDR-----CKMRDLETPHFRWTLTLWF 252
Db 393 SVTHIKTIHQEDELIEIOSDTGTWQYRWGVRLSLGGVGVNGFLSCFSPRYRITLMMMG 452
Qy 253 IWFNSAFSYVGLVL 266
Db 453 VNFMTSFSYGLTV 466

RESULT 9
Q9W3W9 PRELIMINARY; PRT; 709 AA.
ID AC Q9W3W9
RA Adamantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amats M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.N., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sinden-Klamis I., Simpson M.C., Skupski M.P., Smith F.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL; AE003438; AAF46193.1; .
DR FlyBase; FBgn0029896; CG3168.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Transmembrane.
KW Transmembrane.
SQ SEQUENCE 709 AA: 77930 MW: 1B5ADIE9D133AE94 CRC64;

Query Match 23.9%; Score 341.5; DB 5; Length 709;
Best Local Similarity 27.1%; Pred. No. 2e-16;
Matches 83; Conservative 65; Mismatches 123; Indels 35; Gaps 6;

Qy 1 EDVAEAIQFGKFWKLSVLTGLAWMADAMEMMILSLAPQLHCEWRLPSMQVALLTSVVF 60
Db 215 ERAILLCGYKPHY ILLAICGLVSTSEMDVISMFSFILPSAECDLDTNTETKWLNSIIP 274
Qy 61 VGMSSSTLWGNISDQYGRKTKLISVLTLYYGLSFAFVYVSWILVRLGVFGIGV 120
Db 275 IGMVGFALWGLADRLGRQCLLSISVNSVFAFFSFVGYCTFLFCRLLSGVGIGGS 334
Qy 121 -POSVTLAEFLPMKARAKCILLTEVFAIGTVFEVVLAVFVMP-SLG-----WR 168
Db 335 GPVWSYFAEFQPKAKGSMLSFMAAFWTEGMLFVSLAWLIIPRTIGTTPYFTYNSMR 304
Qy 169 WLLILSAVPLLLFAVLCFELPESARYDYLSGNOEKAIATLKRIATENGAPMLGLIISR 220
Db 395 IFLVCSLPSELVGLFLLFELPESPKFLTRGKKDRALALFRGIFVNTKRRDEYVVDL 454
Qy 229 Q-----EDRGMKRDLF-----TPHFRWTLTLWTFNSAFSYVGLVL 266
Db 455 EVDEKLLSNGVKNKYSRMISGMVDHSRALFKSPILRFTIVSI-TINFTHIGYGLIM 513
Qy 267 LTTELF 272
Db 514 WPELF 519

RESULT 10
Q90406 PRELIMINARY; PRT; 724 AA.
ID AC Q90406
RA 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TRANSMEMBRANE TRANSPORTER.
OS Discopyote ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hypnosquales; Pristioraja; Batoidea;
OC Torpediniformes; Narcinoidae; Narcinidae; Discopyge.
OX NCBI_Taxid=7785;

```



QY 167 WRLLILSAVPLLLFAVLCEWLPESARYDVLSGNQEKAIAIATLKRIATEN-----GAP---M 219  
ID 094840 PRELIMINARY; PRT; 683 AA.  
AC 094840;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE P87=TRANSPORTER-LIKE PROTEIN.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93050176; PubMed=1426240;  
RA Jorgensen P.N., Fremerey R.T., Jr., Caron M.G.;  
RA Jorgensen P.N., Fremerey R.T., Jr., Caron M.G.;  
RT "Identification, characterization, and molecular cloning of a novel  
RT transporter-like protein localized to the central nervous system."  
RL FEBS Lett. 312:115-122(1992).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL: S47919; AAB24028.1; -  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF00083; sugar\_tr; 1.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 742 AA; 82589 MW; 58AE0C9BB224DEB8 CRC64;

Query Match 23.7%; Score 338; DB 6; Length 742;  
Best Local Similarity 27.4%; Pred. No. 3.6e-16;  
Matches 86; Conservative 54; Mismatches 126; Indels 48; Gaps 5;

QY 1 EDVAEAIQFGKFWKLSVLTGLAMNADAMMMILSILAPQLHCEWRLPSWOVALLTSVVF 60  
ID 153 EAILRECGHGRFQWTLFVGLGLMADGVFVVGFLPSAEKDMCLSDSNKMGMLGIVY 212  
QY 61 VGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGG- 119  
ID 213 LGMMVGAFLMGGLADRLGRQCLLSLVNSVFAFFSSVQGYGTFLECRLLSGVGIGGS 272  
QY 120 VQSVTLVLAEPKARAKCILLIEVFAIGTVFVWLAFFVMPSLG----- 166  
ID 273 IPIVFSYSEFLAQKREHLSWLCFMWIGGVYAAAWAIIPIHYGWSFQMGSAQPHS 332  
QY 167 WRLLILSAVPLLLFAVLCEWLPESARYDVLSGNQEKAIAIATLKRIATEN-----GAP---M 219  
ID 333 WRVFLVCAFPVFAIGALTTPQSPRPFLENGKHDEAWMLKQVHDTNMRAGHPERV 392  
QY 220 PLGLKLIISRODR-----GKMRDLFTPHFRWTTLLLWF 252  
ID 393 SVTHKTHQDELLIEIOSDTGANTQVGRVRLSLGGQVGMNLFSCFGPEYRITLMMMG 452  
QY 253 IWFSAFYSYGLV 266  
ID 453 VWFMSFSYGLTV 466

RESULT 14  
O94840  
ID 094840 PRELIMINARY; PRT; 683 AA.  
AC 094840;  
DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE KIAA0735.  
GN KIAA0735.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=BRIN;  
RX MEDLINE=99087487; PubMed=9872452;  
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,  
RA Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XI.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro."  
RL DNA Res. 5:277-286(1998).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL: AB018278; BAA34455.1; -  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF00083; sugar\_tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 683 AA; 77444 MW; B981812089BF09E2 CRC64;

Query Match 23.2%; Score 331; DB 4; Length 683;  
Best Local Similarity 27.4%; Pred. No. 1e-15;  
Matches 86; Conservative 60; Mismatches 120; Indels 48; Gaps 6;

QY 1 EDVAEAIQFGKFWKLSVLTGLAMNADAMMMILSILAPQLHCEWRLPSWOVALLTSVVF 60  
ID 96 ETIMDECGHGRFQWTLFVGLGLMADGVFVVGFLPSAEKDMCLSDSNKMGMLGIVY 155  
QY 61 VGMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVRLGLVGFIGG- 119  
ID 156 LGMMVGAFLMGGLADRLGRQCLLSLVNSVFAFFSSVQGYGTFLECRLLSGVGIGGA 215  
QY 120 VQSVTLVLAEPKARAKCILLIEVFAIGTVFVWLAFFVMPSLG----- 166  
ID 216 LPIVFAVSEFLSKRGEHLSWLCFMWIGGVYAAAWAIIPIHYGWSFQMGSAQPHS 275  
QY 167 WRLLILSAVPLLLFAVLCEWLPESARYDVLSGNQEKAIAIATLKRIATEN-----GAPMPLG 222  
ID 276 WRVFLVCAFPVFAIGALTTPQSPRPFLENGKHDEAWMLKQVHDTNMRAGHPERV 335  
QY 223 KL-----IISROEDRG-----KMRDLFTP-----HFRWTTLLLWF 252  
ID 336 TVSNIKTPKQDEFTIQSSTGTQVRLVRFKTIQKQVNDNALCYVMGPMYRMTLLAV 395  
QY 253 IWFSAFYSYGLV 266  
ID 396 VWFMAFYSYGLTV 409  
RESULT 15  
O9HYD9  
ID 09HYD9 PRELIMINARY; PRT; 455 AA.  
AC 09HYD9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE PROBABLE MFS TRANSPORTER.  
GN PA3467.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 12:59:50 ; Search time 24.74 Seconds  
(without alignments)  
1640.750 Million cell updates/sec

Title: US-09-911-667A-2  
 Perfect score: 548  
 Sequence: 1 NEEDLFQRLPVPVKFRTG.....MYGRGMHGAGVTKSNSGSOE 548

Scoring table: OLIGO  
Gapop 60.0 . Gapext 60.0

Searched: 522463 seqs, 74073290 residues

```

Word size : 0
Total number of hits satisfying chosen parameters: 192911

```

```
Minimum DB seq length: 100
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

```
Database :
A_Geneseq_1101.*
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	548	100.0		548	21	AAV44533	Human organic cati
2	7	1.3		114	22	AAQ90545	C glutamicum prote
3	7	1.3		135	21	AAQ36095	zea mays protein f
4	7	1.3		137	20	AAV04945	Mycobacterium spec
5	7	1.3		140	20	AAV41321	Human secreted pro
6	7	1.3		140	20	AAV04946	Mycobacterium spec
7	7	1.3		153	22	AAQ82372	S. epidermidis ope
8	7	1.3		162	22	AAQ73460	Human gene 4-encod
9	7	1.3		213	22	AAQ75328	Gene 44 human secr
10	7	1.3		217	22	AAQ82731	S. epidermidis ope
11	7	1.3		230	17	AAQ85954	Eukaryotic cell gr

12	7	1.3	236	21	AAG30331	Arabidopsis thaliana
13	7	1.3	257	21	AA770102	Staphylococcus aureus
14	7	1.3	309	22	AAW43625	Human polypeptide
15	7	1.3	329	21	AAG30330	Arabidopsis thaliana
16	7	1.3	333	21	AAB13689	C. pneumoniae CT52
17	7	1.3	333	22	AAG82898	S. epidermidis ope
18	7	1.3	333	22	AAG83257	Protein encoded by
19	7	1.3	334	20	AAV35268	Amino acid sequenc
20	7	1.3	346	20	AAV21851	Human signal pepti
21	7	1.3	346	21	AAG36309	Arabidopsis thaliana
22	7	1.3	368	16	AAK68961	Xenorhabdus nemato
23	7	1.3	368	20	AAW97810	Xenorhabdus nemato
24	7	1.3	396	18	AAW34218	Streptomyces hydro
25	7	1.3	396	19	AAW55819	Streptomyces roseo
26	7	1.3	401	21	AAV96284	Sinorhizobium meli
27	7	1.3	420	18	AAW15074	M. leprae gyrase i
28	7	1.3	430	17	AAW03567	Arabidopsis thaliana
29	7	1.3	451	22	AAG90061	C glutamicum prote
30	7	1.3	451	22	AAK76258	Corynebacterium gl
31	7	1.3	461	21	AAG36308	Arabidopsis thaliana
32	7	1.3	467	20	AAW85687	DBX oligosaccharid
33	7	1.3	511	22	AAW41958	Human polypeptide
34	7	1.3	549	22	AAK67471	Amino acid sequenc
35	7	1.3	550	22	AAE06612	Human protein havi
36	7	1.3	550	22	AAE69091	Human organic anio
37	7	1.3	599	22	AAE06591	Human protein havi
38	7	1.3	670	19	AAW29639	Human secreted pro
39	7	1.3	774	19	AAW80411	A secreted protein
40	7	1.3	890	16	AAK77845	Human Rse rPTK H
41	7	1.3	890	21	AAK27663	Human protein PRO8
42	7	1.3	1266	22	AAW40172	Human polypeptide
43	7	1.3	1273	18	AAW15078	M. leprae gyrA pre
44	7	1.3	1297	19	AAW40200	Infected cell prot
45	7	1.3	1298	19	AAW80810	The amino acid seq

## ALIGNMENTS

RESULT	1
XX	AAAY44633
XX	ID AAY44633 standard; Protein; 548 AA.
XX	AC
XX	AAAY44633;
XX	DT
XX	DE
XX	07-APR-2000 (first entry)
XX	Human organic cation transporter-like protein (OCT1p).
XX	Human; organic cation transporter-like protein; OCT1p;
XX	transmembrane; nootropic; neuroprotective; neuroleptic
XX	antiParkinsonian; antidepressant; cellular process; ce
XX	screen; treatment; prevention; diagnosis; neurodegener
XX	Alzheimer; Parkinson's; Huntington; ALS; amyotrophic l
XX	CNS disorder; central nervous system; schizophrenia; d
XX	behavioural; sleep disorder; Alzheimer's; eating disor
XX	Homo sapiens.
XX	OS
XX	Key
XX	Location/Qualifiers
FT	1..85
FT	/label= Cytoplasmic_domain
FT	141..154
FT	/label= Cytoplasmic_domain
FT	199..208
FT	/label= Cytoplasmic_domain
FT	259..314
FT	/label= Cytoplasmic_domain
FT	395..402
FT	/label= Cytoplasmic_domain
FT	448..457
FT	/label= Cytoplasmic_domain
FT	511..548
FT	/label= Cytoplasmic_domain





XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX Corynebacterium glutamicum.  
 OS  
 XX EP1108790-A2.  
 PN 20-JUN-2001.  
 PD 18-DEC-2000; 2000EP-0127688.  
 XX 16-DEC-1999; 95JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX WPI; 2001-376931/40.  
 DR N-PSDB; AAH65764.  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX Claim 17; SEQ ID NO: 4299; 246pp + Sequence Listing; English.  
 PS The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX Sequence 114 AA;  
 SQ

Query Match  
 Best Local Similarity 100.0%; Score 7; DB 22; Length 114;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 155 ISVLWTL 161  
 Db 58 isvltl 64  
 RESULT 3  
 ID AAG36095  
 XX AAG36095 standard; Protein; 135 AA.  
 AC AAG36095;  
 XX 18-OCT-2000 (first entry)  
 DT Zea mays protein fragment SEQ ID NO: 44187.  
 DE Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence; corn.  
 XX Zea mays subsp. mays.  
 OS  
 XX

PN EP1033405-A2.  
 XX 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF 25-FEB-1999; 99US-0121825.  
 XX 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 18-MAY-1999; 99US-0134370.  
 PR 19-MAY-1999; 99US-0134768.  
 PR 20-MAY-1999; 99US-0134941.  
 PR 21-MAY-1999; 99US-0135124.  
 PR 24-MAY-1999; 99US-0135353.  
 PR 25-MAY-1999; 99US-0135629.  
 PR 27-MAY-1999; 99US-0136021.  
 PR 28-MAY-1999; 99US-0136392.  
 PR 01-JUN-1999; 99US-0136782.  
 PR 03-JUN-1999; 99US-0137222.  
 PR 04-JUN-1999; 99US-0137528.  
 PR 07-JUN-1999; 99US-0137502.  
 PR 08-JUN-1999; 99US-0137724.  
 PR 10-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 14-JUN-1999; 99US-0138847.  
 PR 16-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 17-JUN-1999; 99US-0139453.  
 PR 18-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 21-JUN-1999; 99US-0139763.  
 PR 22-JUN-1999; 99US-0139817.  
 PR 23-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 24-JUN-1999; 99US-0140354.  
 PR 28-JUN-1999; 99US-0140695.  
 PR 29-JUN-1999; 99US-0140823.  
 PR 30-JUN-1999; 99US-0140991.  
 PR 01-JUL-1999; 99US-0141287.  
 PR 99US-0141842.

PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147392.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151303.  
PR 01-SEP-1999; 99US-0151438.  
PR 07-SEP-1999; 99US-0151930.  
PR 10-SEP-1999; 99US-0152363.  
PR 13-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 1.3% Score 7; DB 21; Length 135;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 GLVGFGI 187  
| | | | |  
Db 14 glvgfgl 20

## RESULT 4

AAU04945  
ID AAU04945 standard; Protein; 137 AA.  
XX  
AC AAU04945;  
XX  
DT 06-JUL-1999 (first entry)  
XX  
DE Mycobacterium species protein sequence 40D.  
XX  
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;  
KW Hybridisation; detection; vaccine; immunisation; infection.  
XX  
OS Mycobacterium sp.  
XX  
PN WO9909186-A2.  
XX  
PD 25-FEB-1999.  
XX  
PF 14-AUG-1998; 98WO-FR01813.  
PR 11-SEP-1997; 97FR-0011325.  
PR 14-AUG-1997; 97FR-0010404.

```

XX PA (INSP ) INST PASTEUR.
XX PI Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;
XX PI Guigueno A;
XX DR WPI; 1999-181045/15.
XX DR N-PSDB; AAX34198.
XX MYcobacterial DNA vectors containing reporter constructs - for
XX PT identifying coding or promoter sequences involved in
XX PT infection-associated protein expression
XX PS Claim 32; Fig 40D; 309pp; French.
XX CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted
XX CC proteins from various Mycobacterium species microorganisms. The
XX CC encoding nucleotide sequences can be used as primers and probes for
XX CC methods for detecting and identifying mycobacteria, especially belonging
XX CC to the M. tuberculosis complex. The encoded proteins can be used in
XX CC vaccines for immunisation against a bacterial or viral infection.
XX SQ Sequence 137 AA;

Query Match 1.3%; Score 7; DB 20; Length 137;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 GLVLLTT 339
DB 116 givlltt 122
|||||

RESULT 5
AAY41321
ID AAY41321 standard; Protein: 140 AA.
AC AAY41321;
DT 02-DEC-1999 (first entry)
DE Human secreted protein encoded by gene 14 clone HMVBS81.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
XX
XX WO9947540-A1.
XX
XX 23-SEP-1999.
XX
XX 18-MAR-1999; 99WO-US05804.
XX
XX 19-MAR-1998; 98US-0078563.
XX 19-MAR-1998; 98US-0078566.
XX 19-MAR-1998; 98US-0078573.
XX 19-MAR-1998; 98US-0078574.
XX 19-MAR-1998; 98US-0078576.
XX 19-MAR-1998; 98US-0078577.
XX 19-MAR-1998; 98US-0078578.
XX 19-MAR-1998; 98US-0078579.
XX 01-APR-1998; 98US-0078581.
XX 01-APR-1998; 98US-0080312.
XX 01-APR-1998; 98US-0080313.
XX 01-APR-1998; 98US-0080314.

```

```

XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Kuben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DK;
XX PI Wei Y, Endress CA, Duan RD, Kyaw H, Ebner R, Lafleur DW;
XX PI Olsen HS, Shi Y, Moore PA;
XX WPI; 1999-562050/47.
XX DR N-PSDB; AAZ24824.
XX
XX New isolated human genes, useful for diagnosis and treatment of e.g.
XX PT cancers, neurological disorders, immune diseases, inflammation or blood
XX PT disorders -
XX PS Claim 11; Page 366; 484pp; English.
XX
XX This sequence represents a secreted human protein encoded by the gene
XX CC clone detailed in the descriptor line. The gene can be used to generate
XX CC fusion proteins by linking to the gene to a human immunoglobulin Fc
XX CC portion (e.g. AAZ24802) for increasing the stability of the fused
XX CC protein as compared to the human protein only.
XX CC The invention relates to 95 novel genes and their fragments (nucleic
XX CC acid sequences: AAZ24811-224907; amino acid sequences AAY4108-Y41404)
XX CC which are useful for preventing, treating or ameliorating medical
XX CC conditions e.g. by protein or gene therapy. Also, pathological
XX CC polypeptides can be diagnosed by determining the amount of the new
XX CC polynucleotides in a sample or by determining the presence of mutations in
XX CC the new polynucleotides. Specific uses are described for each of the 95
XX CC polynucleotides, based on which tissues they are most highly expressed in
XX CC (see AAZ24811 for described uses).
XX SQ Sequence 140 AA;

Query Match 1.3%; Score 7; DB 20; Length 140;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 LSVITGL 92
DB 63 lsvltgl 69
|||||

RESULT 6
AAY04946
ID AAY04946 standard; Protein: 140 AA.
AC AAY04946;
XX
XX 06-JUL-1999 (first entry)
XX
XX Mycobacterium species protein sequence 40F.
XX
XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
XX KW hybridisation; detection; vaccine; immunisation; infection.
XX
XX Mycobacterium sp.
XX
XX WO9909186-A2.
XX
XX 25-FEB-1999.
XX
XX 14-AUG-1998; 98WO-FR01813.
XX
XX 11-SEP-1997; 97FR-0011325.
XX 14-AUG-1997; 97FR-0010404.
XX
XX (INSP ) INST PASTEUR.
XX
XX Gicquel B, Lim EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;
XX PI Guigueno A;
XX WPI; 1999-181045/15.
XX DR

```

DR N-PSDB; AAX34199.  
 XX Mycobacterial DNA vectors containing reporter constructs - for  
 PT identifying coding or promoter sequences involved in  
 PT infection-associated protein expression  
 XX  
 PS Claim 32; Fig 40F; 309pp; French.  
 XX  
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted  
 CC proteins from various Mycobacterium species microorganisms. The  
 CC encoding nucleotide sequences can be used as primers and probes for  
 CC methods for detecting and identifying mycobacteria, especially belonging  
 CC to the M. tuberculosis complex. The encoded proteins can be used in  
 CC vaccines for immunisation against a bacterial or viral infection.  
 XX  
 SQ Sequence 140 AA;

Query Match 1.3%; Score 7; DB 20; Length 140;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 GLVLLTT 339  
 Db 119 glvlltt 125  
 |||||

RESULT 7  
 AAG82372  
 ID AAG82372 standard; Protein; 153 AA.  
 XX  
 AC AAG82372;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:1838.  
 XX  
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
 KW vaccination; endocarditis.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 XX  
 PF 09-NOV-2000; 2000WO-US30782.  
 XX  
 PR 09-NOV-1999; 99US-0164258.  
 XX  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI Kimmery WJ;  
 XX  
 DR WPI; 2001-316495/33.  
 DR N-PSDB; AAH53222.  
 XX  
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 PT useful for vaccinating against infections, e.g. endocarditis.  
 XX  
 PS Claim 18; Page 506; 2188pp; English.  
 XX  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to

CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 153 AA;

Query Match 1.3%; Score 7; DB 22; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 280 RIATENG 286  
 Db 23 riateng 29  
 |||||

RESULT 8  
 AAG73460  
 ID AAG73460 standard; Protein; 162 AA.  
 XX  
 AC AAG73460;  
 XX

DT 10-AUG-2001 (first entry)  
 XX

DE Human gene 4-encoded secreted protein fragment, SEQ ID NO:235.  
 XX

KW Human; secreted protein; proliferative disorder; cancer; chromosome 11;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angioygenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder; tumour;  
 KW endocrine disorder; infection; wound healing; vulnery;  
 KW cell culture; chemotaxis; food additive;  
 KW binding partner identification.

XX Homo sapiens.  
 OS

XX WO200134628-A1.  
 PN

XX 17-MAY-2001.  
 PD

XX 08-NOV-2000; 2000WO-US30653.  
 PF

XX 12-NOV-1999; 99US-0164735.  
 PR

XX 27-JUL-2000; 2000US-0221193.  
 PR

XX (HUMA-) HUMAN GENOME SCI INC.  
 PA

XX Ruben SM, Komatsoulis GA, Birse CE, Ni J, Moore PA;  
 PI

XX WPI; 2001-329066/34.  
 DR

XX Nucleic acids encoding 35 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy -  
 PT

PS Disclosure; Page 14; 604pp; English.  
 XX

CC AAH32522-AAH32627 represent cDNAs corresponding to 35 human secreted  
 CC protein genes, and AAG73346-AAG73448 represent the proteins they encode.  
 CC AAG73449-AAG73519 represent human secreted protein fragments. The genes  
 CC and their corresponding secreted proteins are useful for preventing,  
 CC treating or ameliorating medical conditions, e.g., by protein or gene  
 CC therapy. Pathological conditions can be diagnosed by determining the

amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 52 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein fragment referred to in the disclosure of the invention.

Sequence 162 AA;

Query Match 1.3%; Score 7; DB 22; Length 162;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 FSPCSLL 416  
Db 41 fscsll 47

RESULT 9  
ID AAB75328  
XX AAB75328 standard; Protein; 213 AA.  
AC AAB75328;  
XX  
DT 03-APR-2001 (first entry)  
XX  
DE Gene 44 human secreted protein homologous amino acid sequence #147.  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic; neurotropic;  
KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnary; autoimmune disease; hyperproliferative disorder; cancer;  
KW cardiovascular disorder; cerebrovascular disorder; infection;  
KW nervous system disorder; ocular disorder; chemotaxis; food additive;  
secreted protein.  
XX  
OS Homo sapiens.  
XX  
XX WO200077021-A1.  
PN  
XX  
XX 21-DEC-2000.  
XX  
XX 01-JUN-2000; 2000WO-US15135.  
PF  
XX  
XX 11-JUN-1999; 99US-0138632.  
PR  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX (ROSE/) ROSEN C A.  
XX  
XX Rosen CA, Ruben SM, Komatsoulis GA;  
PI  
XX WPI; 2001-071257/08.  
DR  
XX

Nucleic acid molecules encoding human secreted proteins, used in preventing, treating or ameliorating a disorder, e.g. Alzheimer's and

PT Parkinson's diseases and cancers -  
XX  
PS Disclosure; Page 96; 530pp; English.  
XX

This invention relates to polynucleotide sequences AAF63789 - AAF63836 which encode human secreted proteins AAB75260 - AAB75287. Included in the invention are protein sequences AAB75288 - AAB75341 which are fragments of the secreted proteins and amino acid sequences with which these fragments share homology. Examples of the activities of the proteins and polynucleotides and the activities of their agonists and antagonists include, immunosuppressive; antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; and vulnary activity. The protein and polynucleotide sequences, their agonists and antagonists may be useful for treating, preventing and diagnosing diseases and disorders such as autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities. Included in the invention are oligonucleotides AAF63780 - AAF63788 and peptide AAB75239 which are used in the identification and characterisation of the DNA and protein sequences of the invention.

Sequence 213 AA;

Query Match 1.3%; Score 7; DB 22; Length 213;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 LSVLTGL 92  
Db 137 lsvltgl 143

RESULT 10  
AAG82731  
ID AAG82731 standard; Protein; 217 AA.  
XX  
AC AAG82731;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2556.  
XX  
KW Staphylococcus epidermidis SK1 strain; infection; diagnosis;  
KW vaccination; endocarditis.  
XX  
OS Staphylococcus epidermidis.  
XX  
XX WO200134809-A2.  
PN  
XX  
XX 17-MAY-2001.  
PD  
XX  
XX 09-NOV-2000; 2000WO-US30782.  
PF  
XX  
XX 09-NOV-1999; 99US-0164258.  
PR  
XX  
XX (GLAX ) GLAXO GROUP LTD.  
PA  
XX  
XX Kimmerly WJ;  
PI  
XX  
XX WPI; 2001-316495/33.  
DR  
XX N-PSDB; AAB53581.  
XX

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
 useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 18; Page 670; 2188pp; English.  
 XX  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG81120, from *Staphylococcus epidermidis*.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC *S. epidermidis* polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX

Sequence 217 AA;

Query Match 1.3%; Score 7; DB 22; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 280 RIATENG 286  
 Db 203 riateng 209  
 |||||

RESULT 11

AAH95954  
 ID AAR95954 standard; Protein: 230 AA.

AC AAR95954;

DT 24-AUG-1996 (first entry)

DE Eukaryotic cell growth inhibiting factor.

KW Eukaryotic cell growth inhibiting factor; ageing; cancer; infection;  
 gene therapy; *Schizosaccharomyces pombe*; fission yeast.

OS Homo sapiens.

PN WO9617933-A2.

PD 13-JUN-1996.

PF 05-DEC-1995; 95WO-JP02488.

PR 02-JUN-1995; 95JP-0136252.

PR 09-DEC-1994; 94JP-0306602.

PR 16-MAR-1995; 95JP-0057716.

PA (TAKE ) TAKEDA CHEM IND LTD.

PI Igarashi K, Sasada R, Takeyama M;

DR WPI; 1996-287178/29.

DR N-PSDB; AAT30564.

XX DNA encoding a eukaryotic cell growth inhibiting factor, and its  
 prod. - useful as anticancer and infection treatment

PS Claim 20; Page 50-51; 73pp; English.

XX A eukaryotic cell growth inhibiting factor (AAR95954) was identified  
 CC as the product of cDNA clone pTB1848 (AAT30564). This clone was  
 CC isolated by incorporating cDNA from aged normal human diploid fibroblast  
 CC MKC-5 cells into vector pTB1589 under control of the nmt1 promoter,  
 CC transforming of *Schizosaccharomyces pombe* cells and screening for  
 CC transformants showing cDNA-dependent growth inhibition. Other cDNA  
 CC clones coding for eukaryotic cell growth inhibiting factors were  
 CC also obtd. (see also AAR95951-57). These proteins are useful as  
 CC anticancer agents and infection remedies. They inhibit nucleic  
 CC acid synthesis in target cells. They can also be used to terminate  
 CC the cell cycle of cultured cells at a given time point.  
 XX

Sequence 230 AA;

Query Match 1.3%; Score 7; DB 17; Length 230;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 71 EDAVEAI 77

Db 122 edaveai 128  
 |||||

RESULT 12

AAG30331

ID AAG30331 standard; Protein: 236 AA.

AC AAG30331;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 36241.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 hybridisation assay; genetic mapping; gene expression control; promoter;  
 termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 24-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0155659.  
PR 29-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.



PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 1.3%; Score 7; DB 21; Length 236;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 EDAAVEAI 77  
 DB 27 edaveai 33

RESULT 13  
 AAY70102  
 ID AAY70102 standard; Protein: 257 AA.

XX AC AAY70102;  
 XX DT 05-JUN-2000 (first entry)  
 XX DE Staphylococcal enterotoxin A.

XX KW Superantigen toxin; SAg; Staphylococcal enterotoxin A; SEA; cytostatic;  
 KW antibacterial; vaccine; MHC class II receptor; T-cell antigen receptor;  
 KW antibody; toxoid; staphylococcal/streptococcal toxin; diagnosis;  
 KW treatment; superantigen-associated bacterial infection.

XX OS Staphylococcus sp.

XX FH Key Location/Qualifiers  
 FT Peptide 2..25  
 FT /label= Leader\_peptide  
 FT Protein 26..253  
 FT /label= Mature\_Staphylococcal\_enterotoxin\_A  
 FT /note= "Includes transcription start site residue, Met"  
 FT Misc-difference 42  
 FT /note= "Encoded by TTG"  
 FT Misc-difference 125  
 FT /note= "Encoded by CCA"

XX WO200009154-A1.  
 XX PD 24-FEB-2000.

XX PF 13-AUG-1998; 98WO-US16766.  
 XX PD 13-AUG-1998; 98WO-US16766.  
 XX PA (REED-) REED ARMY INST RES WALTER.

XX PI Ulrich RG, Olson MA, Bavari S;  
 XX DR WPI: 2000-224177/19.  
 XX DR N-PSDB; AAZ51105.

XX PT Nucleic acid encoding superantigen toxin useful as a vaccine and for  
 XX diagnosis of superantigen-associated bacterial infections -  
 XX PS Claim 7; Page 72-73; 118pp; English.

XX CC The present amino acid sequence is the Staphylococcal enterotoxin A  
 CC (SEA), a bacterial superantigen toxin (SAg). The coding region  
 CC of the SAg toxin when altered by site directed mutagenesis, results  
 CC in disruption of binding of the toxin to both the MHC class II or T-cell  
 CC antigen receptor. SEA has antibacterial and cytostatic activity. This  
 CC sequence is useful for the production of SEA vaccines and specific

CC antibodies. This vaccine overcomes the disadvantages of the chemically  
 CC inactivated toxoids and is designed to protect individuals against one  
 CC or several related staphylococcal and streptococcal toxins. It is used  
 CC for the diagnosis and treatment or amelioration of superantigen.  
 CC associated bacterial infections.

XX SQ Sequence 257 AA;

Query Match 1.3%; Score 7; DB 21; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 427 TLLLFIA 433  
 DB 7 tlllfia 13

RESULT 14  
 AAM43625  
 ID AAM43625 standard; Protein: 309 AA.

XX AC AAM43625;  
 XX DT 22-OCT-2001 (first entry)  
 XX DE Human polypeptide SEQ ID NO 303.

XX KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;  
 KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
 KW fungicide; ophthalmological; cytotoxic; immunosuppressive; neurotropic;  
 KW neuroprotective; antiallergic; hepatotropic; antidiabetic;  
 KW antiinflammatory; antitumor; vulnerary; anticonvulsant; antibacterial;  
 KW antiparasitic; cardiant; gene therapy; cancer; immune disorder;  
 KW cardiovascular disorder; neurological disease; infection; human.

XX OS Homo sapiens.

XX PN WO200155308-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01309.  
 XX PR 31-JAN-2000; 2000US-0179065.  
 XX PR 04-FEB-2000; 2000US-0180628.  
 XX PR 24-FEB-2000; 2000US-0184664.  
 XX PR 02-MAR-2000; 2000US-0186350.  
 XX PR 16-MAR-2000; 2000US-0189874.  
 XX PR 17-MAR-2000; 2000US-0190076.  
 XX PR 18-APR-2000; 2000US-0198123.  
 XX PR 19-MAY-2000; 2000US-0205515.  
 XX PR 07-JUN-2000; 2000US-0209467.  
 XX PR 28-JUN-2000; 2000US-0214886.  
 XX PR 30-JUN-2000; 2000US-0215135.  
 XX PR 07-JUL-2000; 2000US-0216647.  
 XX PR 07-JUL-2000; 2000US-0216880.  
 XX PR 11-JUL-2000; 2000US-0217487.  
 XX PR 11-JUL-2000; 2000US-0217496.  
 XX PR 14-JUL-2000; 2000US-0218290.  
 XX PR 26-JUL-2000; 2000US-0220963.  
 XX PR 26-JUL-2000; 2000US-0220964.  
 XX PR 14-AUG-2000; 2000US-0224518.  
 XX PR 14-AUG-2000; 2000US-0224519.  
 XX PR 14-AUG-2000; 2000US-0225213.  
 XX PR 14-AUG-2000; 2000US-0225214.  
 XX PR 14-AUG-2000; 2000US-0225266.  
 XX PR 14-AUG-2000; 2000US-0225267.  
 XX PR 14-AUG-2000; 2000US-0225268.  
 XX PR 14-AUG-2000; 2000US-0225270.  
 XX PR 14-AUG-2000; 2000US-0225447.  
 XX PR 14-AUG-2000; 2000US-0225757.  
 XX PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251866.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX  
XX WPI; 2001-488781/53.  
DR N-PSDB; AA163931.  
XX  
PT New isolated nucleic acids and polypeptides, useful for diagnosing,  
PT treating and/or preventing human diseases and disorders -  
XX  
XX Claim 11; SEQ ID NO 303; 664pp + Sequence Listing; English.  
PS  
XX The invention relates to human polynucleotides (AA163803-AA164012) and  
CC the encoded proteins (AA1634497-AA163660) useful for preventing, treating  
CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
CC genes were isolated from a range of human tissues disclosed in the  
CC specification. The nucleic acids, proteins, antibodies and (ant)agonists  
CC are useful in the diagnosis, treatment and prevention of: (a) cancer,  
CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 309 AA;

Query Match 1.3%; Score 7; DB 22; Length 309;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 LLIEVF 217 PR 18-JUN-1999; 99US-0139454.  
Db 264 llievfw 270 PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

RESULT 15  
AAG30330  
ID AAG30330 standard; Protein; 329 AA.  
AC AAG30330;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 36240.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 08-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.

Job time: 86 sec

```
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149002.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 1.3%; Score 7; DB 21; Length 329;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 71 EDAVEAI 77  
| | | | |  
Db 120 edaveai 126



; FILING DATE: 30-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-236356  
; FILING DATE: 30-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-189274  
; FILING DATE: 11-AUG-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-189273  
; FILING DATE: 11-AUG-1945  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 6-189272  
; FILING DATE: 11-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Resnick, David S.  
; REGISTRATION NUMBER: 34,235  
; REFERENCE/DOCKET NUMBER: 45753  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-523-3400  
; TELEFAX: 617-523-6440  
; INFORMATION FOR SEQ ID NO: 351:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 115 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-513-974B-351

Query Match 1.3%; Score 7; DB 3; Length 115;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 TGLAWMA 96  
|||||||  
Db 15 TGLAWMA 21

RESULT 2  
US-08-486-099-113  
; Sequence 113, Application US/08486099  
; Patent No. 6013263  
; GENERAL INFORMATION:  
; APPLICANT: Bolognesi, Dani P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.  
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
; TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HEPATITIS  
; TITLE OF INVENTION: B VIRUS TRANSMISSION  
; NUMBER OF SEQUENCES: 209  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,099  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.

; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7872-031  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 113:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 257 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-486-099-113

Query Match 1.3%; Score 7; DB 3; Length 257;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 TLLLFIA 433  
|||||||  
Db 7 TLLLFIA 13

RESULT 3  
US-08-360-107A-123  
; Sequence 123, Application US/08360107A  
; Patent No. 6017536  
; GENERAL INFORMATION:  
; APPLICANT: Bolognesi, Dani P.  
; APPLICANT: Matthews, Thomas J.  
; APPLICANT: Wild, Carl T.  
; APPLICANT: Barney, Shawn O.  
; APPLICANT: Lambert, Dennis M.  
; APPLICANT: Petteway, Stephen R.  
; APPLICANT: Langlois, Alphonse J.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
; TITLE OF INVENTION: TRANSMISSION  
; NUMBER OF SEQUENCES: 149  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/360,107A  
; FILING DATE: 20-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7872-013  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 123:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 257 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-360-107A-123

Query Match 1.3%; Score 7; DB 3; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433  
 Db 7 TLLLFIA 13

## RESULT 4

US-08-484-223B-113  
 ; Sequence 113, Application US/08484223B  
 ; Patent No. 6020459  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bolognesi, Dani P.  
 ; APPLICANT: Matthews, Thomas J.  
 ; APPLICANT: Wild, Carl T.  
 ; APPLICANT: Barney, Shawn O.  
 ; APPLICANT: Lambert, Dennis M.  
 ; APPLICANT: Petteway, Stephen R.  
 ; APPLICANT: Langlois, Alphonse J.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
 ; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
 ; NUMBER OF SEQUENCES: 245  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/919,597  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/470,896  
 ; FILING DATE: 06-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7872-029  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 113:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 257 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-484-223B-113

Query Match 1.3%; Score 7; DB 3; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433  
 Db 7 TLLLFIA 13

## RESULT 5

US-08-919-597-113  
 ; Sequence 113, Application US/08919597

Patent No. 6054265  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bolognesi, Dani P.  
 ; APPLICANT: Matthews, Thomas J.  
 ; APPLICANT: Wild, Carl T.  
 ; APPLICANT: Barney, Shawn O.  
 ; APPLICANT: Lambert, Dennis M.  
 ; APPLICANT: Petteway, Stephen R.  
 ; APPLICANT: Langlois, Alphonse J.  
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
 ; TITLE OF INVENTION: OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
 ; NUMBER OF SEQUENCES: 273  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/919,597  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/470,896  
 ; FILING DATE: 06-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Coruzzi, Laura A.  
 ; REGISTRATION NUMBER: 30,742  
 ; REFERENCE/DOCKET NUMBER: 7872-029  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 113:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 257 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-919-597-113

Query Match 1.3%; Score 7; DB 3; Length 257;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433  
 Db 7 TLLLFIA 13

## RESULT 6

US-08-475-668A-113  
 ; Sequence 113, Application US/08475668A  
 ; Patent No. 6060065  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barney, Shawn O.  
 ; APPLICANT: Lambert, Dennis M.  
 ; APPLICANT: Petteway, Stephen R.  
 ; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
 ; TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS  
 ; NUMBER OF SEQUENCES: 211  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICANT: Bolognesi, Dani P.  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-026  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 113:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 257 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-475-668A-113

Query Match 1.3%; Score 7; DB 3; Length 257;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433  
|||||  
DB 7 TLLLFIA 13

RESULT 7  
US-08-485-551A-113  
Sequence 113, Application US/08485551A  
Patent No. 6068973  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE  
FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 211  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICANT: Bolognesi, Dani P.  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-023  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 113:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 257 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-485-551A-113

Query Match 1.3%; Score 7; DB 3; Length 257;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 427 TLLLFIA 433  
|||||  
DB 7 TLLLFIA 13

RESULT 8  
US-08-471-913A-113  
Sequence 113, Application US/08471913A  
Patent No. 6093794  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 214  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICANT: Bolognesi, Dani P.  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-030  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 113:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 257 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown



MOLECULE TYPE: protein  
US-08-471-913A-113

Query Match 1.3% Score 7; DB 3; Length 257;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 TLLLFIA 433  
Db 7 TLLLFIA 13

## RESULT 9

US-08-485-264A-113  
Sequence 113, Application US/08485264A  
Patent No. 6228983  
GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING  
TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION  
NUMBER OF SEQUENCES: 232  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485,264A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-021  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 113:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 257 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: protein  
MOLECULE TYPE: protein  
US-08-485-264A-113

Query Match 1.3% Score 7; DB 4; Length 257;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 427 TLLLFIA 433  
Db 7 TLLLFIA 13

## RESULT 10

US-08-569-168-2  
Sequence 2, Application US/08569168  
Patent No. 5972687  
GENERAL INFORMATION:  
APPLICANT: Smigielski, Adam J.  
APPLICANT: Akhurst, Raymond J.  
TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lowe, Price, Leblanc & Becker  
STREET: 99 Canal Center Plaza, Suite 300  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/569,168  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Price, Robert L.  
REGISTRATION NUMBER: 22,685  
REFERENCE/DOCKET NUMBER: 1451-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-684-1111  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 278 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-569-168-2

Query Match 1.3% Score 7; DB 2; Length 278;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 240 LLLLSAV 246  
Db 190 LLLLSAV 196

## RESULT 11

US-08-569-168-7  
Sequence 7, Application US/08569168  
Patent No. 5972687  
GENERAL INFORMATION:  
APPLICANT: Smigielski, Adam J.  
APPLICANT: Akhurst, Raymond J.  
TITLE OF INVENTION: TOXIN GENE FROM XENORHABDUS NEMATOPHILUS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lowe, Price, Leblanc & Becker  
STREET: 99 Canal Center Plaza, Suite 300  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/569,168  
FILING DATE:

```

; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Price, Robert L.
; REGISTRATION NUMBER: 22,685
; REFERENCE/DOCKET NUMBER: 1451-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-684-1111
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 383 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-569-168-7

```

```

Query Match      1.3%; Score 7; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 240 LLILSAV 246
      |||||
Db 190 LLILSAV 196

```

## RESULT 12

```

; Sequence 2, Application US/08601435
; Patent No. 5759801
; GENERAL INFORMATION:
; APPLICANT:

```

```

; TITLE OF INVENTION: DNA sequence coding for a protein
; TITLE OF INVENTION: of A, thaliana having a delta-5,7 sterol,delta-7
; TITLE OF INVENTION: reductase activity, delta7-Red protein, production
; TITLE OF INVENTION: process, strains of transformed yeasts, uses.
; NUMBER OF SEQUENCES: 31
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/601,435
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: FR 9501723
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9506517
; FILING DATE: 01-JUN-1995
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-601-435-2

```

```

Query Match      1.3%; Score 7; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 425 VLTLLLF 431
      |||||
Db 390 VLTLLLF 396

```

## RESULT 13

```

; US-08-931-047-2

```

```

; Sequence 2, Application US/08931047
; Patent No. 5965417
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: DNA sequence coding for a protein of
; TITLE OF INVENTION: A. thaliana having a delta-5,7 sterol,
; TITLE OF INVENTION: delta-7 reductase activity, delta7-Red
; TITLE OF INVENTION: protein, production process, strains
; TITLE OF INVENTION: of transformed yeasts, uses.
; NUMBER OF SEQUENCES: 31
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,047
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9501723
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9506517
; FILING DATE: 01-JUN-1995
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-931-047-2

```

```

Query Match      1.3%; Score 7; DB 2; Length 430;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 425 VLTLLLF 431
      |||||
Db 390 VLTLLLF 396

```

## RESULT 14

```

; US-08-783-202-2
; Sequence 2, Application US/08783202
; Patent No. 5989881
; GENERAL INFORMATION:
; APPLICANT:

```

```

; TITLE OF INVENTION: DNA sequence coding for a protein of A.
; TITLE OF INVENTION: thaliana having a delta-5,7 sterol,delta-7 reductase
; TITLE OF INVENTION: activity, delta7-Red protein, production process, strains
; TITLE OF INVENTION: of transformed yeasts, uses.
; NUMBER OF SEQUENCES: 31
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/783,202
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9501723
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9506517
; FILING DATE: 01-JUN-1995
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: amino acid

```

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-783-202-2

Query Match 1.3%; Score 7; DB 2; Length 430;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 425 VLTLLLF 431  
Db 390 VLTLLLF 396

RESULT 15  
US-08-445-640-2  
Sequence 2, Application US/08445640  
Patent No. 5709858  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Scadden, David T.  
APPLICANT: Baker, Kevin P.  
APPLICANT: Baron, Will F.  
TITLE OF INVENTION: Protein Tyrosine Kinases  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/445,640  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: 854C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1896  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 890 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-445-640-2

Query Match 1.3%; Score 7; DB 1; Length 890;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 499 LLAALAS 505  
Db 26 LLAALAS 32

Search completed: March 7, 2002, 13:01:37  
Job time: 107 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 12:59:50 ; Search time 19.42 Seconds  
(without alignments)  
2149.519 Million cell updates/sec

Title: US-09-911-667a-2  
Perfect score: 548  
Sequence: 1 NEEDLQRLQLPVVKFRRTG.....MVGGMHGAGVTRNSGSGE 548

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 189327

Minimum DB seq length: 100  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : JIR\_68:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	110	20.1	144	2	T50622	hypothetical prote
2	19	3.5	300	2	S15786	glucose transport
3	19	3.5	529	2	T23190	hypothetical prote
4	9	1.6	331	2	S2163	sucrose specific r
5	9	1.6	331	2	E5878	sucrose specific t
6	8	1.5	236	2	D69048	conserved hypothet
7	8	1.5	293	2	B83783	hypothetical prote
8	8	1.5	359	2	A42462	acetoin catabolism
9	8	1.5	413	1	S34305	nodulation protein
10	8	1.5	413	2	JC5178	probable starvatio
11	8	1.5	425	2	F70608	hypothetical prote
12	8	1.5	472	2	T34748	transmembrane tran
13	8	1.5	826	2	F85510	probable acyl-CoA
14	8	1.5	826	2	F64746	probable membrane
15	8	1.5	1256	2	S60461	gene flightless-1
16	7	1.3	105	2	T18029	hypothetical prote
17	7	1.3	137	2	H70547	hypothetical prote
18	7	1.3	141	1	HATJD	hemoglobin alpha-D
19	7	1.3	141	2	G70144	hypothetical prote
20	7	1.3	145	1	PSKF2U	phospholipase A2 (
21	7	1.3	178	2	D71451	MC0653 homolog PH0
22	7	1.3	178	2	G71195	hypothetical prote
23	7	1.3	180	2	G82858	hypothetical prote
24	7	1.3	190	2	T09136	ADP-ribosylation f
25	7	1.3	192	2	S43563	RO1H10.5 protein -
26	7	1.3	194	1	HSTR1	histone H1 - trout
27	7	1.3	194	2	D72242	conserved hypothet
28	7	1.3	200	2	S67524	calcium-binding pr
29	7	1.3	206	1	HSTRIR	histone H1 - rainb

ALIGNMENTS

RESULT 1

T50622  
hypothetical protein DKF2p761H039.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C:Accession: T50622  
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.  
submitted to the Protein Sequence Database, June 2000  
A:Reference number: 225145  
A:Accession: T50622  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-144 <AAA>  
A:Cross-references: EMBL:AL359592  
A:Experimental source: adult amygdala; clone DKF2p761H039  
C:Genetics:  
A>Note: DKF2p761H039.1

Query Match 20.1% Score 110; DB 2; Length 144;  
Best Local Similarity 100.0%, Pred. No. 5.1e-103;  
Matches 110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 LCFVIFSCSLLLFCVGRNVLTLLIFARAFISGFOAAAYVYTPVPTATRALGLGTC 464

Db 1 LCFVIFSCSLLLFCVGRNVLTLLIFARAFISGFOAAAYVYTPVPTATRALGLGTC 60

QY 465 SCMARVGALITPFIAQVMLESSVYLTAVYSGCCLLAALASCFLPETK 514

Db 61 SCMARVGALITPFIAQVMLESSVYLTAVYSGCCLLAALASCFLPETK 110

RESULT 2

S15786  
glucose transport protein homolog - Caenorhabditis elegans (fragment)

C:Species: Caenorhabditis elegans

C>Date: 06-Jan-1995 #sequence\_revision 05-May-2000 #text\_change 05-May-2000

C:Accession: S15786

R:Craxton, M.; Ainscough, R.; Coulson, A.; Dear, S.; Du, Z.; Durbin, R.; Green, P.; H

Staden, R.; Sulston, J.; Thierry-Mieg, J.; Thomas, K.; Waterston, R.; Wilson, R.

submitted to the EMBL Data Library, May 1991

A:Reference number: S15786

A:Accession: S15786

A:Molecule type: DNA

A:Residues: 1-300 <CRA>

A:Cross-references: EMBL:Z11115

C:Genetics:

A:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3

C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match

3.5% Score 19; DB 2; Length 300;

Best Local Similarity 100.0%; Pred. No. 7.5e-11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 GFGIGGVQSVTLTLYAEFLP 202  
Db 183 GFGIGGVQSVTLTLYAEFLP 201

RESULT 3  
T23190  
hypothetical protein ZK637.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Jun-2000  
C:Accession: T23190  
R:Craxton, M.  
submitted to the EMBL Data Library, April 1993  
A:Reference number: Z19704  
A:Accession: T23190  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-529 <WIL>  
A:Cross-references: EMBL:Z22175; PIDN:CAA80131.1; GSPDB:GN00021; CESP:ZK637.1  
C:Genetics:  
A:Gene: CESP:ZK637.1  
A:Map position: 3  
A:Insertions: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3; 422/2; 472/3  
C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match 3.5%; Score 19; DB 2; Length 529;  
Best Local Similarity 100.0%; Pred. No. 1.2e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 GFGIGGVQSVTLTLYAEFLP 202  
Db 183 GFGIGGVQSVTLTLYAEFLP 201

RESULT 4  
S52163  
sucrose specific repressor - Escherichia coli  
C:Species: Escherichia coli  
C:Date: 07-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 08-Oct-1999  
C:Accession: S52163  
R:Bockmann, J.  
submitted to the EMBL Data Library, September 1994  
A:Reference number: S52160  
A:Accession: S52163  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <HOC>  
A:Cross-references: EMBL:X81461; NID:g608705; PIDN:CAA57220.1; PID:g608709  
C:Superfamily: lac repressor

Query Match 1.6%; Score 9; DB 2; Length 331;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 LCFWLPESA 262  
Db 181 LCFWLPESA 189

Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E85878  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <STO>  
A:Cross-references: GB:AE005174; NID:g12516731; PIDN:AA057489.1; GSPDB:GN00145; UWGF  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z3626

Query Match 1.6%; Score 9; DB 2; Length 331;  
Best Local Similarity 100.0%; Pred. No. 1;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 LCFWLPESA 262  
Db 181 LCFWLPESA 189

RESULT 6  
D69048  
conserved hypothetical protein MTH1364 - Methanobacterium thermoautotrophicum (strain  
C:Species: Methanobacterium thermoautotrophicum  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
C:Accession: D69048  
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge,  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H:  
A:Reference number: A69000; MUID:98037514  
A:Accession: D69048  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Residues: 1-236 <MTH>  
A:Cross-references: GB:AE000899; GB:AE000666; NID:g2622468; PIDN:AAB85841.1; PID:g26  
A:Experimental source: strain Delta H  
C:Genetics:  
A:Gene: MTH1364

Query Match 1.5%; Score 8; DB 2; Length 236;  
Best Local Similarity 100.0%; Pred. No. 7.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 RGLVGFGI 187  
Db 82 RGLVGFGI 89

RESULT 7  
B83783  
hypothetical protein BH1066 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 31-Dec-2000  
C:Accession: B83783  
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.;  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans  
A:Reference number: A83650; MUID:20263314  
A:Accession: B83783  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-293 <STO>  
A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BAB04785.1; GSPDB:  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH1066

Query Match 1.5%; Score 8; DB 2; Length 293;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 LLLFAVLC 255  
|||||  
DB 20 LLLFAVLC 27

RESULT 8  
A42462  
acetoin catabolism protein AcoX - Alcaligenes eutrophus (strain H16)  
C:Species: Alcaligenes eutrophus  
C:Date: 24-Jul-1992 #sequence\_revision 24-Jul-1992 #text\_change 30-Sep-1993  
C:Accession: A42462  
R: Priefert, H.; Hein, S.; Krueger, N.; Zeh, K.; Schmidt, B.; Steinbuechel, A.  
J. Bacteriol. 173, 4056-4071, 1991  
A:Title: Identification and molecular characterization of the Alcaligenes eutrophus H16  
A:Reference number: A42462; MUID:91286190  
A:Accession: A42462  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-359 <PRI>  
A:Cross-references: GB:M66060

Query Match 1.5%; Score 8; DB 2; Length 359;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 LLAALASC 506  
|||||  
DB 49 LLAALASC 56

RESULT 9  
S34305  
nodulation protein nodC - Rhizobium sp.  
C:Species: Rhizobium sp.  
C:Date: 13-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 16-Jul-1999  
C:Accession: S34305  
R: Relix, B.; Perret, X.; Golinowsky, W.; Pueppke, S.G.; Krishnan, H.B.; Broughton, W.J.  
submitted to the EMBL Data Library, June 1993  
A:Description: Lipo-oligosaccharide Nod-factor signals permit rhizobial penetration into  
A:Reference number: S34303  
A:Accession: S34305  
A:Molecule type: DNA  
A:Residues: 1-413 <REL>  
A:Cross-references: EMBL:X73362; NID:g312347; PIDN:CAA51774.1; PID:g312350  
C:Comment: This is one of the proteins, coded by nodulation genes, that are required for  
C:Genetics:  
A:Gene: nodC  
C:Superfamily: nodulation protein nodC  
C:Keywords: nodulation

Query Match 1.5%; Score 8; DB 1; Length 413;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 LSVLTGLA 93  
|||||  
DB 319 LSVLTGLA 326

RESULT 10  
JC5178  
probable starvation-sensing protein A - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 16-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 31-Jan-2000  
C:Accession: JC5178; T42059  
R: Schneider, D.; Bruton, C.J.; Chater, K.F.  
Gene 177, 243-251, 1996

Query Match 1.5%; Score 8; DB 2; Length 413;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 DRLGRKKT 402  
|||||  
DB 70 DRLGRKKT 77

RESULT 12  
T34748  
transmembrane transport protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
C:Accession: T34748  
R: Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1998  
A:Reference number: Z21556  
A:Accession: T34748  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-472 <NUR>  
A:Cross-references: EMBL:AL031184; PIDN:CAA20171.1; GSPDB:GN00070; SCOEDB:SC2A11.02c  
A:Experimental source: strain A3(2)  
C:Genetics:

A:Title: Characterization of spaA, a Streptomyces coelicolor gene homologous to a gen  
A:Reference number: JC5178; MUID:97080529  
A:Accession: JC5178  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-413 <SCH>  
A:Cross-references: EMBL:X94190; NID:g1694907; PIDN:CAA63900.1; PID:e222100; PID:g169  
C:Comment: This protein is involved in an intercellular signalling system.  
C:Genetics:  
A:Gene: spaA  
A:Start codon: GTG  
F:221/Active site: Asp #status predicted  
F:247,273/Active site: Glu #status predicted

Query Match 1.5%; Score 8; DB 2; Length 413;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 VELDDGAA 55  
|||||  
DB 383 VELDDGAA 390

RESULT 11  
F70608  
hypothetical protein Rv1200 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C:Accession: F70608  
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
A:Reference number: A70500; MUID:98295987  
A:Accession: F70608  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-425 <COL>  
A:Cross-references: GB:Z93777; GB:AL123456; NID:g3261726; PIDN:CAB07823.1; PID:g19290  
A:Experimental source: strain H37Rv  
C:Genetics:  
A:Gene: Rv1200  
C:Superfamily: citrate utilization determinant

Query Match 1.5%; Score 8; DB 2; Length 425;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 DRLGRKKT 402  
|||||  
DB 70 DRLGRKKT 77

RESULT 12  
T34748  
transmembrane transport protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
C:Accession: T34748  
R: Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, August 1998  
A:Reference number: Z21556  
A:Accession: T34748  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-472 <NUR>  
A:Cross-references: EMBL:AL031184; PIDN:CAA20171.1; GSPDB:GN00070; SCOEDB:SC2A11.02c  
A:Experimental source: strain A3(2)  
C:Genetics:

A:Gene: SCQEDB:SC2All.02c

C:Superfamily: citrate utilization determinant

Query Match 1.5%; Score 8; DB 2; Length 472;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 395 DRLGRKKT 402

|||||

Db 86 DRLGRKKT 93

RESULT 13

F85510  
probable acyl-CoA dehydrogenase (EC 1.3.99.-) yafH [imported] - Escherichia coli (strain  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C:Accession: F85510  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,  
Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: F85510

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-826 <STO>

A:CROSS-references: GB:AF0051174; NID:g12512989; PIDN:AAG54546.1; GSPDB:GN00145; UWGP:202

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: yafH

C:Keywords: oxidoreductase

Query Match 1.5%; Score 8; DB 2; Length 826;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 MMILSILA 108

|||||

Db 13 MMILSILA 20

RESULT 14

F64746

probable membrane protein b0221 - Escherichia coli

N:Alternate names: hypothetical protein b0221

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 08-Oct-1999

C:Accession: F64746

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: F64746

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-826 <BLAT>

A:CROSS-references: GB:AE000130; GB:U00096; NID:g1786402; PIDN:AAC73325.1; PID:g1786414;

A:Experimental source: strain K-12, substrain MG1655

F:10-26/Domain: transmembrane #status predicted <TM1>

F:33-49/Domain: transmembrane #status predicted <TM2>

F:51-67/Domain: transmembrane #status predicted <TM3>

F:292-308/Domain: transmembrane #status predicted <TM4>

Query Match 1.5%; Score 8; DB 2; Length 826;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 101 MMILSILA 108

Db 13 MMILSILA 20

RESULT 15

S60461

gene flightless-I protein - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 19-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 21-Jul-2000

C:Accession: S60461; T08425

R:de Couet, H.G.; Fong, K.S.K.; Weeds, A.G.; McLaughlin, P.J.; Miklos, G.L.G.

Genetics 141, 1049-1059, 1995

A:Title: Molecular and mutational analysis of a gelsolin-family member encoded by the

A:Reference number: S60461; MUID:96129280

A:Accession: S60461

A:Molecule type: DNA

A:Residues: 1-1256 <DEC>

A:CROSS-references: GB:AF017777; EMBL:U28044; NID:g3004652; PIDN:AAC28407.1; PID:g3030

R:Maloszka, R.; de Couet, H.G.; Miklos, G.L.

Proc. Natl. Acad. Sci. U.S.A. 95, 3731-3736, 1998

A:Title: Data transferability from model organisms to human beings: insights from the

A:Reference number: Z16415; MUID:98188272

A:Accession: T08425

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1256 <NAL>

A:CROSS-references: GB:AF017777; GB:U80043; GB:U28044; NID:g3004652; PIDN:AAC28407.1;

C:Genetics:

A:Gene: fli-I; flightless-I

A:CROSS-references: FlyBase:FBgn000070

A:Introns: 18/3; 1070/2; 1120/3

C:Superfamily: leucine-rich alpha-2-glycoprotein repeat homology; gelsolin repeat hom

F:53-75/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>

F:76-100/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR2>

F:101-123/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>

F:124-147/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>

F:148-170/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>

F:171-193/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR6>

F:220-242/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR7>

F:243-265/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR8>

F:266-288/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR9>

F:497-830/Domain: gelsolin repeat homology <GEL1>

F:892-1250/Domain: gelsolin repeat homology <GEL2>

Query Match 1.5%; Score 8; DB 2; Length 1256;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 LFQLRQLP 12

|||||

Db 184 LFQLRQLP 191

Search completed: March 7, 2002, 13:02:04

Job time: 134 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 13:00:50 ; Search time 13.29 seconds  
(without alignments)  
1511.838 Million cell updates/sec

Title: US-09-911-667A-2  
Perfect score: 548  
Sequence: 1 MEEDIFQLRLPVVKFRRTG.....MVGRGMHGAGVTRNSNGSUE 548

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 88480

Minimum DB seq length: 100

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	3.5	529	1	P10638 caenorhabdi
2	9	1.6	331	1	P40715 escherichia
3	8	1.5	359	1	P27748 alceguen
4	8	1.5	413	1	P50357 rhizobium s
5	8	1.5	1256	1	P10062 phenodon p
6	7	1.3	141	1	P00606 bangarus mu
7	7	1.3	145	1	P02254 salmo trutt
8	7	1.3	194	1	P02254 salmo trutt
9	7	1.3	202	1	P80363 orconectes
10	7	1.3	206	1	P06350 oncorhynch
11	7	1.3	229	1	P08363 bluetongue
12	7	1.3	229	1	P04684 bluetongue
13	7	1.3	229	1	P04685 bluetongue
14	7	1.3	229	1	P04686 bluetongue
15	7	1.3	229	1	P13841 bluetongue
16	7	1.3	229	1	P23705 bluetongue
17	7	1.3	229	1	P04687 bluetongue
18	7	1.3	242	1	Q00825 odontella s
19	7	1.3	255	1	P51751 rhodospiril
20	7	1.3	257	1	P13163 staphylococ
21	7	1.3	258	1	Q9yc01 aeropyrum p
22	7	1.3	292	1	Q8xz18 drosophila
23	7	1.3	306	1	P96680 bacillus su
24	7	1.3	315	1	P34212 caenorhabdi
25	7	1.3	319	1	P05186 pseudomonas
26	7	1.3	323	1	P06010 rhodopseudo
27	7	1.3	324	1	Q9ktb6 vibrio chol
28	7	1.3	347	1	P30993 mus musculu
29	7	1.3	356	1	Q15365 homo sapien
30	7	1.3	356	1	P09048 oryctolagus
31	7	1.3	373	1	O67266 aquifex aeo
32	7	1.3	381	1	P57445 klebsiella
33	7	1.3	415	1	P39664 synechococc

RESULT 1

ID	YOUNG_CAEEL	STANDARD:	PRT:	529 AA.
DT	01-APR-1993 (Rel. 25, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III.			
GN	ZK637.1.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RC	MEDLINE=92168156; PubMed=1538779;			
RA	Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,			
RA	Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A.,			
RA	Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,			
RA	Ainscough R., Waterston R.;			
RT	*The C. elegans genome sequencing project: a beginning.*;			
RL	Nature 356:37-41(1992).			
RN	[2]			
RP	REVIEWS.			
RC	STRAIN=BRISTOL N2;			
RA	Durbin R.;			
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.			
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).			
CC	-!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z11115; CAA77460.1; -			
DR	EMBL; Z22175; CAA77460.1; JOINED.			
DR	EMBL; Z22175; CAA80131.1; -			
DR	EMBL; Z11115; CAA80131.1; JOINED.			
DR	PIR; S15786; S15786			
DR	WormPep; ZK637.1; CE066638			
DR	InterPro; IPR003662; sub_transportr.			
DR	Pfam; PF00083; sugar_tr; 1.			
DR	PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.			
DR	PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.			
KW	Hypothetical protein; Transmembrane; Transport.			
FT	TRANSMEM 86 106 POTENTIAL.			
FT	TRANSMEM 122 142 POTENTIAL.			
FT	TRANSMEM 158 178 POTENTIAL.			
FT	TRANSMEM 238 258 POTENTIAL.			
FT	TRANSMEM 320 340 POTENTIAL.			
FT	TRANSMEM 373 393 POTENTIAL.			

P24151 rhizobium l  
Q51955 pseudomonas  
P28364 eupiotoc oc  
P15663 influenza a  
P48019 oryza sativ  
P47533 mycoplasma  
O9pjp6 chlamydia m  
P21186 mumps virus  
P42359 streptococc  
P45758 escherichia  
P53118 saccharomyc  
P41639 pinus thunb

ALIGNMENTS



FT TRANSMEM 411 431 POTENTIAL.  
 FT TRANSMEM 482 502 POTENTIAL.  
 SQ SEQUENCE 529 AA; 58317 MW; 8D2FF4CBA15ECD2D CRC64;

Query Match 3.5%; Score 19; DB 1; Length 529;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-10;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 GFGIGGVPOSVTLYAEFLP 202  
 DB 183 GFGIGGVPOSVTLYAEFLP 201

RESULT 2  
 ID CSCR\_ECOLI STANDARD; PRT; 331 AA.  
 AC P40715;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE SUCROSE OPERON REPRESSOR (CSC OPERON REGULATORY PROTEIN).  
 GN CSCR.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EC3132;  
 RA Bockmann J.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: REPRESSOR FOR THE CSC OPERON. BINDS D-FRUCTOSE AS  
 AN INDUCER (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE LACI FAMILY OF TRANSCRIPTIONAL  
 REGULATORS.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

EMBL; X81461; CAA57220.1; -.  
 DR HSP; P03023; LCED.  
 DR InterPro; IPR000843; HTH\_LacI.  
 DR InterPro; IPR001761; Peripla\_BP\_like.  
 DR Pfam; PF00356; lacI; 1.  
 DR Pfam; PF00532; Peripla\_BP\_like; 1.  
 DR PRINTS; PRO0036; HTHLACI.  
 DR SMART; SM00354; HTH\_LACI; 1.  
 DR PROSITE; PS00356; HTH\_LACI\_FAMILY; 1.  
 DR Transcription regulation; DNA-binding; Repressor.  
 KW DNA\_BIND 4 23 H-T-H MOTIF (POTENTIAL).  
 FT SEQUENCE 331 AA; 36470 MW; BF9AEA07C10B431F CRC64;  
 SQ

Query Match 1.6%; Score 9; DB 1; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 0.75;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 LCFWLPESA 262  
 DB 181 LCFWLPESA 189

RESULT 3  
 ID ACOX\_ALCEU STANDARD; PRT; 359 AA.  
 AC P27748;  
 DT 01-AUG-1992 (Rel. 23, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 01-DEC-1992 (Rel. 24, Last annotation update)  
 DE ACETOIN CATABOLISM PROTEIN X.  
 GN ACOX.  
 OS Alcaligenes eutrophus (Ralstonia eutropha).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=510;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HI6;  
 RX MEDLINE=91286190; PubMed=2061286;  
 RA Priefert H., Hein S., Krueger N., Zeh K., Schmidt B.,  
 RA Steinbuechel A.;  
 RT Identification and molecular characterization of the Alcaligenes  
 eutrophus HI6 aco operon genes involved in acetoin catabolism.\*;  
 RL J. Bacteriol. 173:4056-4071(1991).  
 CC -!- FUNCTION: ESSENTIAL FOR ACETOIN CATABOLISM.  
 CC -!- PATHWAY: ACETOIN CATABOLISM.  
 CC -!- SUBCELLULAR LOCATION: MEMBRANE (POTENTIAL).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

EMBL; M66060; AAA21947.1; -.  
 DR Membrane.  
 KW  
 SQ SEQUENCE 359 AA; 37934 MW; 6825322F5B8C50E5 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 7.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 499 LLAALASC 506  
 DB 49 LLAALASC 56

RESULT 4  
 ID NODC\_RHISN STANDARD; PRT; 413 AA.  
 AC P50357;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE N-ACETYLGUCOSAMINYLTRANSFERASE (EC 2.4.1.-) (MODULATION PROTEIN C).  
 GN NODC OR Y4HG.  
 OS Rhizobium sp. (strain NGR234).  
 OG Plasmid sym pNGR234a.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97305956; PubMed=9163424;  
 RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
 RA Perret X.;  
 RT Molecular basis of symbiosis between Rhizobium and legumes.\*;  
 RL Nature 387:394-401(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95075295; PubMed=7984092;  
 RA Relic B., Perret X., Estrada-Garcia M.T., Kopicinska J., Golinowski W.,  
 RA Krishnan H.B., Pueppke S.G., Broughton W.J.;  
 RT "Nod factors of Rhizobium are a key to the legume door.\*";  
 RL Mol. Microbiol. 13:171-178(1994).  
 CC -!- FUNCTION: INVOLVED IN THE SYNTHESIS OF NOD FACTOR, A SULFATED  
 N-ACYL-BETA-1,4-TETRASACCHARIDE OF N-ACETYLGUCOSAMINE WHICH

[4]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY; TISSUE=Embryo;  
RX MEDLINE=201960012; PubMed=10731138;  
RR Rubin G.M., Hong L., Brokstein P., Evans-Holm M., Frise E.,  
RA Stapleton M., Harvey D.A.;  
RT "A Drosophila complementary DNA resource.\*";  
RL Science 287:2222-2224(2000).  
  
[5]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RR Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.O., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Benos P.V., Bereman B.P., Bhargava S., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A.C., Chandra I.,  
RA Cherry J.M., Casway S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dudan-Kocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbalwala C.,  
RA Jalali M., Kalush F., Karpen G.H., Kravitz S., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Sanders R.D.C., Scheeter F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Swirskas R., Tector C., Turner R., Venter E., Wissen A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zheng G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.\*";  
RL Science 287:2185-2195(2000).  
  
CC -|- FUNCTION: MAY PLAY A KEY ROLE IN EMBRYONIC CELLULARIZATION BY  
CC INTERACTING WITH BOTH THE CYTOSKELETON AND OTHER CELLULAR  
CC COMPONENTS. ALTERNATIVELY, IT MAY PLAY A STRUCTURAL ROLE IN  
CC INDIRECT FLIGHT MUSCLE. VITAL FOR EMBRYONIC DEVELOPMENT.  
CC -|- TISSUE SPECIFICITY: FOUND IN OVARIES, LARVAL FAT BODIES, BRAIN AND  
CC ADULT THORAX.  
CC -|- DEVELOPMENTAL STAGE: EXPRESSED BOTH MATERNALLY AND ZYGOTICALLY.  
CC -|- DOMAIN: CONSISTS OF A LEUCINE-RICH AMINO TERMINAL HALF, WHICH IS  
CC LIKELY TO BE INVOLVED IN PROTEIN-PROTEIN INTERACTION, AND A  
CC CARBOXYTERMINAL HALF WHICH HAS HIGH SEQUENCE SIMILARITY TO  
CC GELSOLIN AND IS THEREFORE LIKELY TO BE INVOLVED IN ACTIN-BINDING.  
CC -|- SIMILARITY: BELONGS TO THE VILLIN/GELSOLIN FAMILY.  
CC -|- SIMILARITY: CONTAINS 15 LEUCINE-RICH REPEATS (LRR).  
CC -|- SIMILARITY: CONTAINS 4 GELSOLIN-LIKE REPEATS.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions in its use by non-profit institutions as long as its content is not modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement.shtml>). You may send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC AND D. HB A IS A TETRAMER OF TWO ALPHA-A AND TWO BETA-1, HB A' IS  
CC A TETRAMER OF TWO ALPHA-A AND TWO BETA-2, HB D IS A TETRAMER OF  
CC TWO ALPHA-D AND TWO BETA-2.  
CC MISCELLANEOUS: SPHENODON'S HBS HAVE PROPERTIES NOT FOUND IN OTHER  
CC REPTILES: POOR COOPERATIVITY, HIGH AFFINITY FOR OXYGEN, SMALL BOHR  
CC AND HALDANE EFFECTS, APPRECIABLE PHOSPHATE EFFECTS (THOSE  
CC PROPERTIES ARE ALSO FOUND IN THE HBS OF PRIMITIVE URODELE AND  
CC CAECILIAN AMPHIBIANS).  
CC -1- SIMILARITY: HIGH DEGREE OF SIMILARITY WITH ALPHA D SEQUENCE FROM  
CC BIRDS AND TURTLE.  
CC PIR: S01137; HATJD.  
CC HSSP: P01958; 2MHB.  
CC InterPro: IPR002338; Alpha\_haem.  
CC InterPro: IPR000971; Globin.  
CC Pfam: PF00042; globin; 1.  
CC PRINTS: PR00612; ALPHAHAE1.  
CC PROSITE: PS01033; GLOBIN; 1.  
CC Heme: Oxygen transport; Respiratory protein; Erythrocyte.  
CC FT METAL 58 IRON (HEME DISTAL LIGAND).  
CC FT METAL 87 IRON (HEME PROXIMAL LIGAND).  
CC SQ SEQUENCE 141 AA; 16272 MW; F5B8E633C9F9AA1 CRC64;  
  
Query Match 1.3%; Score 7; DB 1; Length 141;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 223 FEVVLAV 229  
Db 105 FEVVLAV 111  
|||||  
  
RESULT 7  
ID PA20\_BUNMU STANDARD; PRT; 145 AA.  
AC P00606;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE PHOSPHOLIPASE A2 PRECURSOR (EC 3.1.1.4) (PHOSPHATIDYLCHOLINE 2-  
DE ACYLHYDROLASE).  
OS Bungarus multicinctus (Many-banded krait).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;  
OC Elapidae; Bungarinae; Bungarus.  
OX NCBI\_TaxID=8616;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Venom gland;  
RX MEDLINE=90356415; PubMed=2388841;  
RA Danse J.M.;  
RT "Nucleotide sequence encoding for non-toxic phospholipase-A2 from  
RT Bungarus multicinctus.";  
RL Nucleic Acids Res. 18:4608-4608(1990).  
RN [2]  
RP SEQUENCE OF 28-145.  
RC TISSUE=Venom;  
RX MEDLINE=81168081; PubMed=7217037;  
RA Kondo K., Toda H., Narita K.;  
RT "Amino acid sequence of phospholipase A from Bungarus multicinctus  
RT venom.";  
RL J. Biochem. 89:37-47(1981).  
CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE  
CC 2-ACYL GROUPS IN 3-SN-PHOSPHOGLYCERIDES.  
CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERYL-  
CC PHOSPHOCHOLINE + A FATTY ACID ANION.  
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified; and this statement is not removed. Usage by and for commercial

DR EMBL; U01182; AAC03566.1; -;  
DR EMBL; AF017777; AAC28407.1; -;  
DR EMBL; AF132184; AAD34772.1; -;  
DR EMBL; AF003568; AAF50830.1; ALT\_SEQ.  
DR HSSP; P02640; 2VIL.  
DR FlyBase; Fgn0000709; f111.  
DR InterPro: IPR001974; Gelsolin.  
DR InterPro: IPR001611; LRR.  
DR InterPro: IPR003592; LRR\_out.  
DR InterPro: IPR003591; LRR\_typ.  
DR Pfam; PF00626; Gelsolin; 4.  
DR Pfam; PF00560; LRR; 12.  
DR PRINTS; PR00019; LEURICHRPT.  
DR PRINTS; PR00597; GELSOLIN.  
DR SMART; SM00262; GEL; 6.  
DR SMART; SM00370; LRR; 3.  
DR SMART; SM00369; LRR\_TYP; 2.  
KW Developmental protein; Repeat; Leucine-rich repeat; Polymorphism.  
FT REPEAT 4 28  
FT REPEAT 29 51  
FT REPEAT 52 74  
FT REPEAT 75 99  
FT REPEAT 100 122  
FT REPEAT 124 145  
FT REPEAT 147 169  
FT REPEAT 171 192  
FT REPEAT 218 241  
FT REPEAT 243 264  
FT REPEAT 265 287  
FT REPEAT 289 312  
FT REPEAT 313 335  
FT REPEAT 336 358  
FT REPEAT 360 381  
FT REPEAT 499 557  
FT REPEAT 746 789  
FT REPEAT 1064 1102  
FT REPEAT 1165 1206  
FT VARIANT 601 601  
FT CONFLICT 1067 1070  
FT CONFLICT 1068 1068  
SQ SEQUENCE 1256 AA; 143681 MW; CF0056EFAA88DB92 CRC64;  
  
Query Match 1.5%; Score 8; DB 1; Length 1256;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 5 LFQLRPL 12  
Db 184 LFQLRPL 191  
|||||  
  
RESULT 6  
ID HBAD\_SPHPU STANDARD; PRT; 141 AA.  
AC P10062;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-JUN-1994 (Rel. 29, Last annotation update)  
DE HEMOGLOBIN ALPHA-D CHAIN.  
OS Sphenodon punctatus (Hatteria) (Tuatara).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosauria; Rhynchocephalia; Sphenodontidae; Sphenodon.  
OX NCBI\_TaxID=8508;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=89105321; PubMed=3214555;  
RA Abbasi A., Wells R.M.G., Brittain T., Braunitzer G.;  
RT "Primary structure of the hemoglobins from Sphenodon (Sphenodon  
RT punctatus, Tuatara, Rhynchocephalia). Evidence for the expression of  
RT alpha D-gene.";  
RL Biol. Chem. Hoppe-Seyler 369:755-764(1988).  
CC -1- SUBUNIT: THERE ARE THREE FORMS OF HEMOGLOBIN IN SPHENODON: A, A'

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X53406; CAA37482.1; -.
DR PIR: S10981; PSKF20.
DR HSSP: P15445; I43F.
DR InterPro: IPR001211; P1P_A2.
DR Pfam: PF00068; Phoslip; 1.
DR PRINTS: PR00389; PHPLIPASEA2.
DR ProDom: PD000303; PLP_A2; 1.
DR SMART: SM00085; PA2c; 1.
DR PROSITE: PS00118; PA2_HIS; 1.
DR PROSITE: PS00119; PA2_ASP; 1.
KW Hydrolase; Lipid degradation; Calcium; Multigene family; Signal;
KW Venom.
FT SIGNAL 1 27
FT CHAIN 28 145 PHOSPHOLIPASE A2.
FT ACT_SITE 73 73 BY SIMILARITY.
FT ACT_SITE 119 119 BY SIMILARITY.
FT ACT_SITE 38 97 BY SIMILARITY.
FT DISULFID 52 144 BY SIMILARITY.
FT DISULFID 54 70 BY SIMILARITY.
FT DISULFID 69 125 BY SIMILARITY.
FT DISULFID 76 118 BY SIMILARITY.
FT DISULFID 86 111 BY SIMILARITY.
FT DISULFID 104 116 BY SIMILARITY.
FT CA_BIND 74 74 BY SIMILARITY.
FT SEQUENCE 145 AA; 15593 MW; F7959376589967CA CRC64;

Query Match 1.3%; Score 7; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 LLILSAV 246
DB 6 LLILSAV 12

RESULT 8
HL_SALTR STANDARD; PRT; 194 AA.
AC P02254;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H1.
OS Salmo trutta (Brown trout).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8032;
RN [1]
RP SEQUENCE
RX MEDLINE=78023898; PubMed=913397;
RA McLeod A.R., Wong N.C.W., Dixon G.H.;
RT "The amino-acid sequence of trout-testis histone H1.";
RL Eur. J. Biochem. 78:281-291(1977).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR PIR: A02583; HSTR1.
DR HSSP: P08287; ICHC.
DR InterPro: IPR001386; Linker_histone.
DR Pfam: PF00538; linker_histone; 1.
DR SMART: SM00526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation; Phosphorylation.
FT MOD_RES 1 1 ACETYLATION (IN 90% OF THE CHAINS).
FT MOD_RES 145 145 PHOSPHORYLATION.
FT MOD_RES 161 161 PHOSPHORYLATION.
FT MOD_RES 182 182 PHOSPHORYLATION.

us-09-911-667a-2.oligo.rsp
```

```
FT DOMAIN 27 100 GLOBULAR.
FT VARIANT 35 35 A -> I (IN MINOR COMPONENT).
SQ SEQUENCE 194 AA; 19408 MW; D288F9F44AF9BE7E CRC64;

Query Match 1.3%; Score 7; DB 1; Length 194;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 354 KKAWEAK 360
DB 100 KKAWEAK 106

RESULT 9
CC23_ORCLI STANDARD; PRT; 202 AA.
ID CC23_ORCLI
AC P80363;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CRUSTACEAN CALCIUM-BINDING PROTEIN 23 (CCBP-23 PROTEIN).
OS Orconectes limosus (Spinycheek crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidea; Cambaridae; Orconectes.
OX NCBI_TaxID=28379;
RN [1]
RP SEQUENCE.
RX TISSUE=Abdominal muscle;
RX MEDLINE=95154352; PubMed=7851448;
RA Sauter A., Staudenmann W., Hughes G.J., Heizmann C.W.;
RT "A novel EF-hand Ca(2+)-binding protein from abdominal muscle of
RT crustaceans with similarity to calyphosine from dog thyroideae.";
RL Eur. J. Biochem. 227:97-101(1995).
CC -1- FUNCTION: POSSIBLY ACTS AS A REGULATORY PROTEIN AND NOT AS A
CC CALCIUM BUFFER OR TRANSPORT PROTEIN.
CC -1- SUBUNIT: MONOMER OR DISULFIDE-LINKED DIMERS.
CC -1- TISSUE SPECIFICITY: STRIATED MUSCLE AND BRAIN.
CC -1- MISCELLANEOUS: THIS PROTEIN BINDS CALCIUM.
CC -1- SIMILARITY: CONTAINS 4 EF-HAND CALCIUM-BINDING DOMAINS.
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; efhand; 4.
DR SMART: SM00054; EFh; 3.
DR PROSITE: PS00018; EF_HAND; FALSE_NEG.
KW Calcium-binding; Repeat; Acetylation.
FT MOD_RES 1 1 ACETYLATION.
FT DOMAIN 46 57 ANCESTRAL CALCIUM SITE 1 (POTENTIAL).
FT CA_BIND 82 93 EF-HAND 2 (POTENTIAL).
FT CA_BIND 118 129 EF-HAND 3 (POTENTIAL).
FT DOMAIN 161 174 ANCESTRAL CALCIUM SITE 4 (POTENTIAL).
FT UNSURE 35 35 OR 1.
FT UNSURE 36 36 OR 1.
SQ SEQUENCE 202 AA; 22848 MW; 191400E4BA8DADD CRC64;

Query Match 1.3%; Score 7; DB 1; Length 202;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 RKKAVEA 359
DB 106 RKKAVEA 112

RESULT 10
HL_ONCMY STANDARD; PRT; 206 AA.
ID HL_ONCMY
AC P06350;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HISTONE H1.
```

```

OS Oncoerhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
ON NCBI_TaxID=8022;
RX SEQUENCE FROM N.A.
RP MEDLINE=85264847; PubMed=6443128;
RA Mezquita J., Connor W., Winkfein R.J., Dixon G.H.;
RT "An H1 histone gene from rainbow trout (Salmo gairdneri).";
RL J. Mol. Evol. 21:209-219(1985).
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02624; CAB37646.1; -
DR PIR; A02584; HSTR1R.
DR HSP; P08287; IGHC.
DR InterPro; IPR001386; Linker_histone.
DR Pfam; PF00538; linker_histone; 1.
DR SMART; SMU0526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Acetylation.
FT INIT_MET 0 0
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT DOMAIN 27 100 GLOBULAR.
FT SEQUENCE 206 AA; 20672 MW; 72C440798066716C CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 206;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 354 KRAVEAK 360
DB 100 KRAVEAK 106
ID VP8_BTIV10 STANDARD; PRT; 229 AA.
AC P08363;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN P8 (NONSTRUCTURAL PROTEIN NS3) [CONTAINS:
DE NONSTRUCTURAL PROTEIN NS3A].
GN S10.
OS Bluetongue virus (serotype 10 / isolate USA).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=10900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87085500; PubMed=3025349;
RA Lee J.W., Roy P.;
RT "Nucleotide sequence of a cDNA clone of RNA segment 10 of bluetongue
RT virus (serotype 10).";
RL J. Gen. Virol. 67:2833-2837(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90345726; PubMed=2166648;
RA Roy P., Marshall J.J.A., French T.J.;
RT "Structure of the bluetongue virus genome and its encoded proteins.";
RL Curr. Top. Microbiol. Immunol. 162:43-87(1990).

```

```

CC -!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CC CELLS.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28981; AAA42840.1; -
DR PIR; A29153; P8XR10.
DR PIR; S10543; S10543.
DR InterPro; IPR002565; Orbi_NS3.
DR Pfam; PF01616; Orbi_NS3; 1.
DR ProDom; PD003183; Orbi_NS3; 1.
KW Nonstructural protein.
FT CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.
FT CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.
FT SEQUENCE 229 AA; 25602 MW; 1A23B53198ECB684 CRC64;
SQ
Query Match 1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 122 VALLTSV 128
DB 130 VALLTSV 136
ID VP8_BTIV11 STANDARD; PRT; 229 AA.
AC Q04684;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN P8 (NONSTRUCTURAL PROTEIN NS3) [CONTAINS:
DE NONSTRUCTURAL PROTEIN NS3A].
GN S10.
OS Bluetongue virus (serotype 11 / isolate USA).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=33716;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295710; PubMed=1318624;
RA Huang G.-Y., Yang Y.-Y., Chiou J.-F., Li J.K.-K.;
RT "Sequence conservation among the cognate nonstructural NS3/3A protein
RT genes of six bluetongue viruses.";
RL Virus Res. 23:151-161(1992).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CC CELLS.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L08631; AAA42835.1; -
DR InterPro; IPR002565; Orbi_NS3.
DR Pfam; PF01616; Orbi_NS3; 1.
DR ProDom; PD003183; Orbi_NS3; 1.
KW Nonstructural protein.
FT CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.
FT CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.
FT SEQUENCE 229 AA; 25374 MW; 22695FC231DED61 CRC64;
SQ

```

```

Query Match      1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 122 VALLTSV 128
    |||||
Db 130 VALLTSV 136

RESULT 13
VP8_BTIV13
ID VP8_BTIV13 STANDARD; PRT; 229 AA.
AC Q04685;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN P8 (NONSTRUCTURAL PROTEIN NS3) [CONTAINS:
DE NONSTRUCTURAL PROTEIN NS3A].
GN S10.
OS Bluetongue virus (serotype 13 / isolate USA).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=33717;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295710; PubMed=1318624;
RA Huang G.-Y., Yang Y.-Y., Chiou J.-F., Li J.K.-K.;
RT "Sequence conservation among the cognate nonstructural NS3/3A protein
RT genes of six bluetongue viruses.";
RL Virus Res. 23:151-161(1992).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CC CELLS.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L08629; AAA42836.1; -
CC InterPro; IPR002565; Orbi_NS3.
CC Pfam; PF01616; Orbi_NS3; 1.
CC ProDom; PD003183; Orbi_NS3; 1.
CC Nonstructural protein.
CC CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.
CC CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.
CC SEQUENCE 229 AA; 25507 MW; 3FDFDIA3138335B8 CRC64;

Query Match      1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 122 VALLTSV 128
    |||||
Db 130 VALLTSV 136

RESULT 14
VP8_BTIV17
ID VP8_BTIV17 STANDARD; PRT; 229 AA.
AC Q04686;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN P8 (NONSTRUCTURAL PROTEIN NS3) [CONTAINS:
DE NONSTRUCTURAL PROTEIN NS3A].
GN S10.
OS Bluetongue virus (serotype 17 / isolate USA).

```

```

OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=33718;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295710; PubMed=1318624;
RA Huang G.-Y., Yang Y.-Y., Chiou J.-F., Li J.K.-K.;
RT "Sequence conservation among the cognate nonstructural NS3/3A protein
RT genes of six bluetongue viruses.";
RL Virus Res. 23:151-161(1992).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CC CELLS.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L08630; AAA42837.1; -
CC InterPro; IPR002565; Orbi_NS3.
CC Pfam; PF01616; Orbi_NS3; 1.
CC ProDom; PD003183; Orbi_NS3; 1.
CC Nonstructural protein.
CC CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.
CC CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.
CC SEQUENCE 229 AA; 25499 MW; 3D99853D39A56CBA CRC64;

Query Match      1.3%; Score 7; DB 1; Length 229;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 122 VALLTSV 128
    |||||
Db 130 VALLTSV 136

RESULT 15
VP8_BTIV1A
ID VP8_BTIV1A STANDARD; PRT; 229 AA.
AC P13841; Q65733;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NONSTRUCTURAL PROTEIN P8 (NONSTRUCTURAL PROTEIN NS3) [CONTAINS:
DE NONSTRUCTURAL PROTEIN NS3A].
GN S10.
OS Bluetongue virus (serotype 1 / isolate Australia).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
OX NCBI_TaxID=10904;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88187687; PubMed=2833571;
RA Gould A.R.;
RT "Nucleotide sequence of the Australian bluetongue virus serotype 1
RT RNA segment 10.";
RL J. Gen. Virol. 69:945-949(1988).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE RELEASE OF VIRIONS FROM INFECTED
CC CELLS.
CC -!- SIMILARITY: BELONGS TO THE REOVIRUSES NS3 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00253; BAA00184.1; -

```

DR EMBL; D00253; BAA00185.1; -.  
DR PIR; A28600; P8XRAU.  
DR InterPro; IPR002565; Orbi\_NS3.  
DR Pfam; PF01616; Orbi\_NS3; 1.  
DR ProDom; PD003183; Orbi\_NS3; 1.  
KW Nonstructural protein.  
FT CHAIN 1 229 NONSTRUCTURAL PROTEIN NS3.  
FT CHAIN 14 229 NONSTRUCTURAL PROTEIN NS3A.  
SQ SEQUENCE 229 AA; 25501 MW; 3132BDCEB6C4325 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 229;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 VALLTSV 128  
|  
Db 130 VALITSV 136

Search completed: March 7, 2002, 13:02:58  
Job time: 128 sec

Result No.	Score	Query Match	Length	DB	ID	Description
1	110	20.1	144	4	Q9NPW5	Q9npw5 homo sapien
2	98	17.9	548	11	Q9Z217	Q9z217 rattus norv
3	13	2.4	497	5	Q9W1D4	Q9w1d4 drosophila
4	8	1.5	205	8	Q78791	Q78791 osmerus mor
5	8	1.5	236	1	Q27417	Q27417 methanobact
6	8	1.5	293	2	Q9KD25	Q9kd25 bacillus ha
7	8	1.5	231	2	Q95726	P95726 streptomyce
8	8	1.5	413	2	Q05301	Q05301 mycobacteri
9	8	1.5	425	2	O05301	O53254 rhizobium t
10	8	1.5	452	2	O05301	O53254 rhizobium t
11	8	1.5	472	2	O86563	O86563 streptomyce
12	8	1.5	814	2	P71283	P71283 escherichia
13	8	1.5	1019	10	P94078	P94078 arabidopsis
14	8	1.5	1060	5	Q917Q3	Q917q3 drosophila
15	7	1.3	101	11	Q9CZK2	Q9czk2 mus musculu
16	7	1.3	103	10	Q9ZRN3	Q9zrn3 galega oxie
17	7	1.3	103	12	Q9QNY8	Q9qny8 human immun
18	7	1.3	105	12	Q98577	Q98577 paramecium
19	7	1.3	107	12	Q9WL65	Q9wl65 human immun
20	7	1.3	121	12	Q9Q7U5	Q9q7u5 human immun

RESULT 2  
Q9Z2I7  
ID Q9Z2I7 PRELIMINARY; PRT; 548 AA.  
AC Q9Z2I7;  
DT 01-MAY-1966, (TREM3Lrel. 10, Created)



DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE SV2 RELATED PROTEIN.  
 GN SVOP.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=99019745; PubMed=9801366;  
 RA Janz R., Hofmann K., Sudhof T.C.;  
 RT "SVOP, an evolutionarily conserved synaptic vesicle protein, suggests  
 RT novel transport functions of synaptic vesicles.";  
 RL J. Neurosci. 18:9269-9281(1998).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
 DR EMBL: AF060173; AAC78627.1; -;  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF00083; sugar\_tr; 1.  
 KW Transmembrane.  
 SQ SEQUENCE 548 AA; 60804 MW; 19AD8475B7579496 CRC64;  
  
 Query Match 17.9%; Score 98; DB 11; Length 548;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-89;  
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 251 FAVLCFWLPESARYDVLGNSQEKATIKRIATENGAMPGLKLIISROEDRGKMRDLPT 310  
 DB 251 FAVLCFWLPESARYDVLGNSQEKATIKRIATENGAMPGLKLIISROEDRGKMRDLPT 310  
 QY 311 PFRWTTLLWFWNSAFSYGLVLLTTELFQAGDVC 348  
 DB 311 PFRWTTLLWFWNSAFSYGLVLLTTELFQAGDVC 348  
  
 RESULT 3  
 Q9WID4 PRELIMINARY; PRT; 497 AA.  
 ID Q9WID4  
 AC Q9WID4;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE CG4324 PROTEIN.  
 GN CG4324.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Balowin D.,  
 RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
 DR EMBL: AE003462; AAF47135.1; -;  
 DR FlyBase: FBgn0034956; CG4324.  
 DR InterPro: IPR003662; sub\_transporter.  
 DR Pfam: PF00083; sugar\_tr; 1.  
 KW Transmembrane.  
 SQ SEQUENCE 497 AA; 55360 MW; 678F44A3286B7BD4 CRC64;  
  
 Query Match 2.4%; Score 13; DB 5; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 0.00043;  
 Matches 1; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 190 VPQSVTLYAERLP 202  
 DB 181 VPQSVTLYAERLP 193  
  
 RESULT 4  
 Q78791 PRELIMINARY; PRT; 205 AA.  
 ID Q78791  
 AC Q78791;  
 DT 01-NOV-1998 (TReMBLrel. 08, Created)  
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)  
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
 DE ATP SYNTHASE A CHAIN (EC 3.6.1.34) (FRAGMENT).  
 OS Osmerus mordax (Rainbow smelt).  
 OG MitoChondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Osmeriformes; Osmeridae; Osmerus.  
 OX NCBI\_TaxID=8014;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pigeon D., Dodson J.J., Bernatchez L.;  
 RT "A mtDNA analysis of spatio-temporal distribution of two genetically  
 RT distinct sympatric larval populations of rainbow smelt (Osmerus  
 RT mordax) in the middle estuary of the St. Lawrence River, Quebec,  
 RT Canada.";  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A  
 CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE  
 CC (BY SIMILARITY).  
 CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC  
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE  
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)  
 CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: TO THE ATPASE A CHAIN FAMILY.  
 DR EMBL: AF034748; AAC24026.1; -;  
 DR InterPro: IPR000568; ATP\_synt\_A.

DR Pfam: PF00119; ATP-synt\_A; 1.  
 DR PRINTS: PR00123; ATPASEA.  
 DR PROSITE: PS00449; ATPASE\_A; 1.  
 KW C(0): Hydrogen ion transport; Mitochondrion; Transmembrane.  
 FT NON\_TER 1  
 SQ SEQUENCE 205 AA; 22598 MW; AE66B1DAB77A984C CRC64;

Query Match 1.5%; Score 8; DB 8; Length 205;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 122 VALLTSVV 129  
 Db 167 VALLTSVV 174  
 |||||

RESULT 5  
 O27417 PRELIMINARY; PRT; 236 AA.  
 AC 027417;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE CONSERVED PROTEIN.  
 GN MTH1364.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
 OC Methanothermobacter.  
 OX NCBI\_TaxID=145262;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DELTA H;  
 RX MEDLINE=98037514; PubMed=9371463;  
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., DuBois J.,  
 RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,  
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,  
 RA Spadafora R., Vicar R., Wang Y., Wierzbowski J., Gibson R.,  
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,  
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,  
 RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;  
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
 RT deltaH: functional analysis and comparative genomics.";  
 RL J. Bacteriol. 179:7135-7155(1997).  
 DR EMBL: AE000899; AAB85841.1; -;  
 DR InterPro: IPR003748; DUF169.  
 DR Pfam: PF02596; DUF169; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 236 AA; 26177 MW; 81B09F4EB67E6BFA CRC64;

Query Match 1.5%; Score 8; DB 1; Length 236;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 180 RGLVGFGI 187  
 Db 82 RGLVGFGI 89  
 |||||

RESULT 6  
 Q9KDD5 PRELIMINARY; PRT; 293 AA.  
 AC Q9KDD5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE BH1066 PROTEIN.  
 GN BH1066.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=86665;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL: AP001510; BAB04785.1; -;  
 DR InterPro: IPR000515; BPD\_transp.  
 DR Pfam: PF00528; BPD\_transp; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 293 AA; 33010 MW; A2B8A16E6217BD56 CRC64;

Query Match 1.5%; Score 8; DB 2; Length 293;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 LLLFAVLC 255  
 Db 20 LLLFAVLC 27  
 |||||

RESULT 7  
 P95726 PRELIMINARY; PRT; 413 AA.  
 ID P95726;  
 AC P95726;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE SPAA (PROBABLE STARVATION SENSING PROTEIN).  
 GN SPAA.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97080529; PubMed=8921874;  
 RA Schneider D., Bruton C.J., Chater K.F.;  
 RT "Characterization of spaA, a Streptomyces coelicolor gene homologous  
 RT to a gene involved in sensing starvation in Escherichia coli.";  
 RL Gene 1177:243-251(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: X94190; CAA63900.1; -;  
 DR EMBL: AL450350; CAC16962.1; -;  
 DR InterPro: IPR001354; MR\_MLE.  
 DR Pfam: PF01188; MR\_MLE; 1.  
 DR PROSITE: PS00908; MR\_MLE\_1; 1.  
 SQ SEQUENCE 413 AA; 45206 MW; 7EDC6183BC8DE36F CRC64;

```

Query Match      1.5%; Score 8; DB 2; Length 413;
Best Local Similarity 100.0%; Pred. No. 35;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 VELDGA 55
DB 383 VELDGA 390
|||||

RESULT 8
Q05301 PRELIMINARY; PRT; 425 AA.
AC Q05301;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 44.6 KDA PROTEIN.
GN RV1200 OR MTC1364.12
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.F., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala K.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seager K., Skellon S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DE EMBL: Z93777; CAB07823.1; -.
DR TubercuList; RV1200; -.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr_1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Complete proteome; Hypothetical protein; Transmembrane.
SQ SEQUENCE 425 AA; 44575 MW; 67B44D03F0EA7256 CRC64;

Query Match      1.5%; Score 8; DB 2; Length 425;
Best Local Similarity 100.0%; Pred. No. 36;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 DRLGRKKT 402
DB 70 DRLGRKKT 77
|||||

RESULT 9
Q53254 PRELIMINARY; PRT; 452 AA.
AC Q53254;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NODA, NODB, NOD C, AND NODD1 GENES.
GN NODC.
OS Rhizobium tropici.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=398;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=CFN299;
RX MEDLINE=97084572; PubMed=8930915;
RA Debell F., Plazanet C., Roche P., Pujol C., Savagnac A., Rosenberg C.,
RA Prome J.C., Denariet J.;
RA "The Noda proteins of Rhizobium meliloti and Rhizobium tropici specify
RT the N-acylation of nod factors by different fatty acids.";
RL Mol. Microbiol. 22:303-314(1996).
DR EMBL: X98514; CAA67139.1; -.
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
SQ SEQUENCE 452 AA; 50019 MW; 2FA3683DCE27A603 CRC64;

Query Match      1.5%; Score 8; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 38;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 LSVLTGLA 93
DB 319 LSVLTGLA 326
|||||

RESULT 10
O86563 PRELIMINARY; PRT; 472 AA.
AC O86563;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE TRANSMEMBRANE TRANSPORT PROTEIN.
GN SC2A11.02C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Murphy L., Harris D.;
RA Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RA Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
DR EMBL: AL031184; CAA20171.1; -.
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR003880; Phosphatn_attach.
DR InterPro: IPR000911; Ribosomal_L11.
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr_1.
DR PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR PROSITE; PS00012; PHOSPHATETHEINE; UNKNOWN_1.
DR PROSITE; PS00359; RIBOSOMAL_L11; UNKNOWN_1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 472 AA; 50203 MW; 4A0CBF6D60DB67E7 CRC64;

Query Match      1.5%; Score 8; DB 2; Length 472;

```

Best Local Similarity 100.0%; Pred. No. 40;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 DRLGRKKT 402  
|||||

Db 86 DRLGRKKT 93

## RESULT 11

ID P71283 PRELIMINARY; PRT; 814 AA.  
AC P71283;  
DT 01-FEB-1997 (TReMBLrel. 02, Created)  
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE HYPOTHETICAL.  
GN YAPF.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,  
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,  
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,  
RA Davis R.W.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U70214; AAB08643.1; -;  
DR InterPro: IPR001552; Acyl-CoA\_dh.  
DR Pfam: PF00441; Acyl-CoA\_dh; 1.  
SQ SEQUENCE 814 AA; 89224 MW; AD9E40ACB44CF781 CRC64;

Query Match 1.5%; Score 8; DB 2; Length 814;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 MMILSILA 108  
|||||

Db 1 MMILSILA 8

## RESULT 12

ID P94078 PRELIMINARY; PRT; 1019 AA.  
AC P94078; O96239;  
DT 01-MAY-1997 (TReMBLrel. 03, Created)  
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE ALPHA-MANNOSIDASE.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucots II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=97086699; PubMed=8932388;  
RA Quigley F., Dao P., Cottet A., Mache R.;  
RT "Sequence analysis of an 81 kb contig from Arabidopsis thaliana  
RT chromosome III";  
RL Nucleic Acids Res. 24:4313-4318(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Quigley F., Dao P., Cottet A., Mache R.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN SEQUENCE FROM N.A.  
RP STRAIN=COLUMBIA;  
RX MEDLINE=20277480; PubMed=10819329;

RA Nakamura Y.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC  
RT clones.";  
RL DNA Res. 7:131-135(2000).

DR EMBL; X98130; CAA66821.1; -;  
DR EMBL; Y11767; CAA72432.1; -;  
DR EMBL; AB026648; BAB01735.1; -;  
DR Mendel; 17407; Arath; 2768; 17407  
DR InterPro: IPR000602; Glyco\_hydro\_38.  
DR InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF01074; Glyco\_hydro\_38; 1.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
SQ SEQUENCE 1019 AA; 115219 MW; 4F764F2359F9FBC CRC64;

Query Match 1.5%; Score 8; DB 10; Length 1019;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 LSGNQEKA 274  
|||||

Db 950 LSGNQEKA 957

## RESULT 13

ID Q91703 PRELIMINARY; PRT; 1060 AA.  
AC Q91703;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE FLII PROTEIN.  
GN FLII OR BCDNA:LD21753 OR CGI484.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA April J.F., Aboyani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari P., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dou P.L.E., Downes M., Dugan Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:40:13 ; Search time 42.48 seconds  
(without alignments)  
955.559 Million cell updates/sec

Title: US-09-911-667A-2  
Perfect score: 2846  
Sequence: 1 MEEDLPQLRPLPVVKFRTG.....MVGKGMHGAGVTRNSGSGSQE 548

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues  
Total number of hits satisfying chosen parameters: 522463

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```
Database :
A_Geneseq_1101.*
1: /SID58/gcgdata/geneseq/geneseqp/AA1980.DAT.*
2: /SID58/gcgdata/geneseq/geneseqp/AA1981.DAT.*
3: /SID58/gcgdata/geneseq/geneseqp/AA1982.DAT.*
4: /SID58/gcgdata/geneseq/geneseqp/AA1983.DAT.*
5: /SID58/gcgdata/geneseq/geneseqp/AA1984.DAT.*
6: /SID58/gcgdata/geneseq/geneseqp/AA1985.DAT.*
7: /SID58/gcgdata/geneseq/geneseqp/AA1986.DAT.*
8: /SID58/gcgdata/geneseq/geneseqp/AA1987.DAT.*
9: /SID58/gcgdata/geneseq/geneseqp/AA1988.DAT.*
10: /SID58/gcgdata/geneseq/geneseqp/AA1989.DAT.*
11: /SID58/gcgdata/geneseq/geneseqp/AA1990.DAT.*
12: /SID58/gcgdata/geneseq/geneseqp/AA1991.DAT.*
13: /SID58/gcgdata/geneseq/geneseqp/AA1992.DAT.*
14: /SID58/gcgdata/geneseq/geneseqp/AA1993.DAT.*
15: /SID58/gcgdata/geneseq/geneseqp/AA1994.DAT.*
16: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT.*
17: /SID58/gcgdata/geneseq/geneseqp/AA1996.DAT.*
18: /SID58/gcgdata/geneseq/geneseqp/AA1997.DAT.*
19: /SID58/gcgdata/geneseq/geneseqp/AA1998.DAT.*
20: /SID58/gcgdata/geneseq/geneseqp/AA1999.DAT.*
21: /SID58/gcgdata/geneseq/geneseqp/AA2000.DAT.*
22: /SID58/gcgdata/geneseq/geneseqp/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	2846	100.0	548	21	Human organic cationic amino acid transporter
2	493.5	17.3	439	22	Corynebacterium glutamicum prote
3	493.5	17.3	448	22	C glutamicum prote
4	441	15.5	535	21	Rat liver anion transporter
5	426.5	15.0	556	17	Rat OCT-1 protein
6	424	14.9	554	19	Human liver cell cationic amino acid transporter
7	402	14.1	542	22	hOAT3. Homo sapiens
8	402	14.1	742	21	Human secreted protein
9	398	14.0	542	21	Human cerebral protein
10	396.5	13.9	536	21	Rat cerebral organic anion transporter
11	395	13.9	537	18	Mouse osteoclast transmembrane protein

12	392.5	13..8	540	22	AAB49401	Murine organic anio
13	387	13..6	561	18	AAM44196	Human osteoclast
14	379	13..3	607	21	AAB12131	Hydrophobic domain
15	374.5	13..2	550	21	AAV44278	Human organic anion
16	374.5	13..2	550	22	AAB47271	HOAT1. Homo sapi
17	374.5	13..2	563	20	AAN88489	Human organic anio
18	370.5	13..0	538	22	AAB47273	HOAT2B. Homo sapi
19	370	13..0	546	22	AAB47272	HOAT2A. Homo sapi
20	367	12..9	548	21	AAB08823	A human organic an
21	361.5	12..7	551	20	AAN88488	Rat organic anion
22	360.5	12..7	551	21	AAB08824	A human organic an
23	357.5	12..6	557	20	AAV01650	A protein with cat
24	357.5	12..6	557	21	AAH83929	Human carnitine tr
25	348	12..2	545	22	AAB36553	Murine organic anio
26	332	11..7	557	20	AAH01652	A protein with cat
27	332	11..7	557	21	AAB20580	Mouse OCTN2 amino
28	332	11..7	557	21	AAH63970	Mouse carnitine tr
29	329.5	11..6	564	21	AAB20578	Mouse OCTN3 protei
30	316.5	11..1	431	22	AAG90943	C glutamicum prote
31	311.5	10..9	399	22	AAB76717	Corynebacterium q
32	311.5	10..9	530	21	AAB29626	Cat flea HMT synap
33	302.5	10..6	551	20	AAV01649	A protein with cat
34	301.5	10..6	560	21	AAB43038	Corynebacterium gl
35	297	10..4	446	22	AAB76810	C glutamicum prote
36	297	10..4	475	22	AAG92363	Human protein havi
37	296.5	10..4	550	22	AAE06612	Human organic anio
38	296.5	10..4	550	22	AAB69091	Human bone marrow
39	295	10..4	584	22	AAW00930	Human protein havi
40	294.5	10..3	578	22	AAC06571	C glutamicum prote
41	288.5	10..1	452	22	AAG93094	A protein with cat
42	286.5	10..1	553	20	AAV01651	Mouse OCTN1 amino
43	286.5	10..1	553	21	AAB20579	Plant phosphate tr
44	284	10..0	585	22	AAB47415	C glutamicum prote
45	283.5	10..0	460	22	AAG92391	

ALIGNMENTS

RESULT	1
AAV44633	
ID	AAV44633 standard; Protein: 548 AA.
XX	
AC	AAV44633;
DT	
DE	07-APR-2000 (first entry)
XX	
WW	Human organic cation transporter-like protein (OCTlp).
KW	Human; organic cation transporter-like protein; OCTlp; transporter;
KW	transmembrane; nootropic; neuroprotective; neuroleptic; anticonvulsant;
KW	antiParkinsonian; antidepressant; cellular process; cell proliferation;
KW	screen; treatment; prevention; diagnosis; neurodegenerative disorder;
KW	Alzheimer; Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;
KW	CNS disorder; central nervous system; schizophrenia; depression;
KW	behavioural; sleep disorder; Alzheimer's; eating disorder.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	Domain
FT	/label= Cytoplasmic_domain
FT	141..154
FT	/label= Cytoplasmic_domain
FT	199..208
FT	/label= Cytoplasmic_domain
FT	259..314
FT	/label= Cytoplasmic_domain
FT	395..402
FT	/label= Cytoplasmic_domain
FT	448..457
FT	/label= Cytoplasmic_domain
FT	511..548

## ALIGNMENTS

RESULT	1
AA44633	
ID	AA44633 standard; Protein; 548 AA.
XX	
XX	AA44633;
XX	AC
XX	DT
XX	07-APR-2000 (first entry)
XX	
DE	Human organic cation transporter-like protein (OCT1p).

xx Human: organic cation transporter-like protein; OCT1p: transporter;  
 kw transmembrane; nootropic; neuroprotective; neuroleptic; anticonvulsant;  
 kw antiparkinsonian; antidepressant; cellular process; cell proliferation;  
 kw screen; treatment; prevention; diagnosis; neurodegenerative disorder;  
 kw Alzheimer; Parkinson's; Huntington; ALS; amyotrophic lateral sclerosis;  
 kw CNS disorder; central nervous system; schizophrenia; depression;  
 kw behavioural; sleep disorder; Alzheimer's; eating disorder

FT Domain /label= Cytoplasmic\_domain  
 FT 86..108  
 FT /label= Transmembrane\_domain  
 FT 122..140  
 FT /label= Transmembrane\_domain  
 FT 155..175  
 FT /label= Transmembrane\_domain  
 FT 182..198  
 FT /label= Transmembrane\_domain  
 FT 209..232  
 FT /label= Transmembrane\_domain  
 FT 239..258  
 FT /label= Transmembrane\_domain  
 FT 315..337  
 FT /label= Transmembrane\_domain  
 FT 376..394  
 FT /label= Transmembrane\_domain  
 FT 403..421  
 FT /label= Transmembrane\_domain  
 FT 428..447  
 FT /label= Transmembrane\_domain  
 FT 458..479  
 FT /label= Transmembrane\_domain  
 FT 487..510  
 FT /label= Transmembrane\_domain  
 FT 109..121  
 FT /label= Extracellular\_domain  
 FT /note= "Hydrophilic region useful for antibody  
 production"  
 FT 176..181  
 FT /label= Extracellular\_domain  
 FT /note= "Hydrophilic region useful for antibody  
 production"  
 FT 233..238  
 FT /label= Extracellular\_domain  
 FT /note= "Hydrophilic region useful for antibody  
 production"  
 FT 338..375  
 FT /label= Extracellular\_domain  
 FT /note= "Hydrophilic region useful for antibody  
 production"  
 FT 422..427  
 FT /label= Extracellular\_domain  
 FT /note= "Hydrophilic region useful for antibody  
 production"  
 FT 480..486  
 FT /label= Extracellular\_domain  
 FT /note= "Hydrophilic region useful for antibody  
 production"  
 FT 71..524  
 FT /note= "shows homology to a consensus sequence for  
 sugar and other transporter molecules derived from  
 a hidden Markov model"  
 PN W0200000633-A1.  
 XX  
 XX  
 XX 06-JAN-2000.  
 XX  
 XX 29-JUN-1999; 99WO-US14880.  
 XX  
 XX 30-JUN-1998; 98US-0107932.  
 XX  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX  
 XX Goodearl AJ, Glucksmann MA;  
 XX WPI; 2000-137069/12.  
 DR N-PSDB; AAZ49686.  
 XX  
 XX New nucleic acid encoding human organic cation transporter-like  
 PT protein, used for prevention, treatment and diagnosis of e.g.  
 PT neurological, behavioural or sleep disorders -  
 XX

PS Claim 9; Fig 1; 100pp; English.  
 XX  
 CC The present sequence is a human OCT1p (organic cation  
 CC transporter-like protein), a member of the superfamily of sugar and  
 CC other transporter molecules that have 12 transmembrane domains. The  
 CC The sequence is derived from a human foetal brain cDNA library. The  
 CC protein is highly expressed in brain tissue and has nootropic,  
 CC neuroprotective, neuroleptic, anticonvulsant, antiparkinsonian,  
 CC antidepressant activities. The present sequence is used to  
 CC regulate a variety of cellular processes e.g. cell proliferation,  
 CC differentiation and survival, screen OCT1p modulators and detect mutation  
 CC in OCT1p gene. OCT1p modulators can be used to treat or prevent chronic  
 CC neurodegenerative disorders (e.g. Alzheimer's, Parkinson's,  
 CC Huntington's and ALS), CNS disorders (e.g. schizophrenia, panic,  
 CC depression), behavioural, sleep and eating disorders.  
 XX  
 SQ Sequence 548 AA;

Query Match 100.0%; Score 2846; DB 21; Length 548;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MEEDLFQLRQLPVVKFRRTGESARSEDDTASGEHEVQIEGVHVGLEAVEVDGAAVPKEF 60  
 Db 1 meedlfqlrqlpvvkfrtgesarседdtasgehevqiegvhvgleaveldgaaavpkef 60  
 Qy 61 ANPTDDTFMVEDAVEAIGFGKFWKLSVLTGLAWMADAMENMILSLAPQLHCEWRLPWSW 120  
 Db 61 anptddtfmvedaveaigfgkfwklsv. glawmadamemmilslapqlhcewrpws 120  
 Qy 121 QVALLTSVYFVGMSSSTLWGNISDQYGRKTGLKLSVLWTLYYGLSAPYVSILVLR 180  
 Db 121 qvallsvfyfvgmssstlwgnisdqygrktglklsvltwlyyglssatapyvsilv 180  
 Qy 181 GLVFGIGGVQSVTLYAEFLPMKARAKCILLIEVFVWAGTVFEVVLAVFVWPSLGRWML 240  
 Db 181 glvfgiggvqsvtlyae flpmkarakcillievfvwagtvfevvlavfvmps lgrwml 240  
 Qy 241 LLSAVPLLLFAVLFCFWLPESARYDVLSCNOEKATATLKRIATENGAPMPLGKLLISRQE 300  
 Db 241 llsavplllfavlcfcfwlpesarydvlsgnqekatatlkriatengapmp l gkllisr 300  
 Qy 301 DRGMRDLTFPHFRWTTLLLMFIFSNAFSYVGLVLLTTELFQAGDVGCGISSRKKAVEAK 360  
 Db 301 drgmrldtfphfrwtllllmfifsnafsyvgvlvtte l f qagdv c g i s s r k k a v e a k 360  
 Qy 361 CSLACEYLSEEDYMDLLWTTLSEPPGVLTWIIDRLGRKKTMLALCFVIFSCSLLLIFIC 420  
 Db 361 cslaceylseedy m d l l w t t l s e f p g v l t w i i d r l g r k k t m a l c f v i f s c s l l l i f i c 420  
 Qy 421 VGRNVLTLLFIARAFISGGFOAAVYVTPYPTATRALGLGTCGSMARVGCALITPPIAQ 480  
 Db 421 vgrnvtlllfiarafisggf o a a v y v t p y p t a t r a l g l g t c g s m a r v g c a l i t p p i a q 480  
 Qy 481 VMLESSVYLTAVYSGCCLLAALASCFPIETKGGGLQESSHREMGVGMHAGV 540  
 Db 481 vmlessvy l t a v y s g c c l l a a l a s c f p i e t k g g g l q e s s h r e m g v g m h a g v t 540  
 Qy 541 RNSNGSQE 548  
 Db 541 rnsngsqe 548

RESULT 2  
 AAB76766  
 ID AAB76766 standard; Protein; 439 AA.  
 XX  
 AC AAB76766;  
 XX  
 DT 11-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum MCT protein SEQ ID NO:514.











QY 239 WLLILLSAVPILLFAVLCFWLPESARYDVLSGNOEKAIATLKRIAT-----ENGAPMPLCK 293  
Db 237 wlgitvsiptvflfsswwtpeisirvlvisgksskalkilrrvavfngkkeeerlslee 296  
QY 294 LITSROED-----RGKMRDLF-TPHFRWTTLLWFTWFSNAFSYXGLVLLTTE----- 340  
Db 297 lkinlqkeislakakytasdlfrpmlrmtfclsawfatgfayyslangveefgvnly 356  
QY 341 -----LFOAGDVCGISSRKKKAVEAKCSLACEYLSEEDYMDLLWTLSEFPQVLVTLWIHDR 396  
Db 357 ilqifigvdv-----pakfittislsy 379  
QY 397 LGRKKTALCFVIFSCSL-LLFICVC-RNVLTLLFIARAFISGGFQAAYVVTPEVYPT 454  
Db 380 lgrhttaqaallagaaialtvtlqvtvrlavfkgkclsssfcliytsetlypt 439  
QY 455 ATRALGTCGSMARVAGALITPPIAQMVLSSSVYLTAVYSGCLLAALASCFLPIETKG 514  
Db 440 vtrqtmgvsnlwtvrgsmvslv-kitgevqfpipnliygitaligssaalfip-etln 497  
QY 515 GGLQES 520  
Db 498 qplpet 503  
RESULT 8  
AA94977  
ID - AA94977 standard; Protein: 742 AA.  
AC AA94977;  
XX  
DT 16-JUN-2000 (first entry)  
XX  
DE Human secreted protein clone as180\_1 protein sequence SEQ ID NO:160.  
XX  
KW Human; secreted protein; immunestimulant; immunesuppressant; virucide;  
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;  
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;  
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;  
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;  
KW connective tissue disease; multiple sclerosis; erythematosis;  
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;  
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;  
KW insulin dependent diabetes mellitus; graft-versus-host-disease;  
KW autoimmune inflammatory eye disease; allergy.  
XX  
OS Homo sapiens.  
XX  
PN WO200009552-A1.  
XX  
PD 24-FEB-2000.  
XX  
PF 13-AUG-1999; 99WO-US18298.  
XX  
PR 14-AUG-1998; 98US-0096622.  
PR 17-AUG-1998; 98US-0096815.  
PR 04-SEP-1998; 98US-0099229.  
PR 23-OCT-1998; 98US-0105368.  
PR 08-JAN-1999; 99US-0115234.  
PR 12-FEB-1999; 99US-0119931.  
PR 18-FEB-1999; 99US-0120575.  
PR 30-APR-1999; 99US-0132020.  
PR 11-AUG-1999; 99US-0096622.  
XX  
PA (GBMY ) GENETICS INST INC.  
XX  
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
PI Wong GG, Clark HF, Fechtel K;  
XX  
DR WPT; 2000-205979/18.  
XX  
XX New polynucleotides encoding secreted proteins, which may have e.g.

PT nutritional, chemokine, immune stimulating or suppressing,  
PT hematopoiesis regulating, tissue growth, activin/inhibin  
PT antiinflammatory or tumor inhibition activity  
XX  
PS Claim 169; Page 614-616; 641pp; English.  
XX  
XX AAA16618 to AAA16697 encode the human secreted proteins given in  
CC AAY94980 to AAY94980, isolated from human adult brain, adult thyroid,  
CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,  
CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,  
CC foetal brain, adult thymus, foetal placenta. The polynucleotides and proteins are  
CC and adult bladder, cDNA libraries. The polynucleotides and proteins are  
CC predicted to have biological activities which would make them suitable  
CC for treating, preventing or ameliorating medical conditions in humans  
CC and animals. The polynucleotides can be used as markers for tissues in  
CC which the protein is preferentially expressed, as molecular weight  
CC markers on Southern gels, and as chromosome markers or tags to identify  
CC chromosomes or to map gene positions. The proteins can be used in the  
CC treatment of immune deficiencies and disorders, such as severe combined  
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other  
CC infections. These infections include human immunodeficiency virus (HIV),  
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and  
CC candidiasis. The proteins can be used to treat autoimmune disorders such  
CC as connective tissue disease, multiple sclerosis, systemic lupus  
CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,  
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent  
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and  
CC autoimmune inflammatory eye disease. The proteins can also be used to  
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent  
CC probes for the human secreted protein; from the present invention.  
XX  
XX Sequence 742 AA;  
QY 21 ESARSEDDTASGE-----VOIEGVHVGLEAVEIDLDGAAP----- 57  
Db 85 eghdedddiyegeyqgpraesggkgkermadgaplagvrg-----lsdgcgppggrgea 139  
QY 58 -----KEFANPTDDTFWEDAVEAIGFQWKLVSULTGLANWADAMEMMILSILAPQL 111  
Db 140 qrtkereeiaq-----gyeailrecqhgfrwtlyfvlglalmadgvefvvvgvlpisa 193  
QY 112 HCEWRLPSQWVALLTSVVFVGMSSSTLWGNISDQYGRKTKISVLTWLYGILSAFAP 171  
Db 194 ekdmclsdnskgmlglivlygmvgafllwggjadrlgrrqccllslsvnsvfaiffsvq 253  
QY 172 VYSWILVLRGLVGFIGG-VPOSVTLXAEFLPMKARAKCILLIEVFWAIGTVFEVWLAVF 230  
Db 254 gygtflcrllsgvgigsgipivfsyfeqlaqekrghlswlcmfwmiggyaamaawa 313  
QY 231 VMPSLG-----WRWLLILSAVPLLLFAVLCFWLPESARYDVLSGNOEKALAT 277  
Db 314 ilphygwsfmgsgayqhsrvfvivcafpvsfaigaltqtqpesprfflengkhdeamv 373  
QY 278 LKRIATEN-----GAP---MPLGKLIISRQEDR-----G 303  
Db 374 lkqvhdntmrakghpervfsvthiktihqedelieiqsdgtgtwyqrgvraalslggqv 433  
QY 304 KMRDLFTPHFRWTTLLWFTWFSNAFSYXGLVL----- 336  
Db 434 nlfscfgpeyrritlmmvgvwtmsfsgygtvfwfpmirhlqavdyasrtkvfpgerve 493  
QY 337 -----LTFELFOAGDVCG-----ISSRKKAVEAKCSL-----ACEYLS 369  
Db 494 hvtfntlenqhrvggyfndkfglrlksvsfedsifeecyfedvtsntffnctfin 553  
QY 370 EEDY-----MDLLWT-----TLSEFPQVL 388  
Db 554 tvfyntdlfeykfvnsrlinstflhnkegcpldvgtgagaymyvfvslgtlavlpgni 613

```

QY 389 VTLWIIDLGRKKTMALCFVIFSCSLLLFICVGRN---VLTLLLFARAFISGGFOAAAY 445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 614 vsallmdkigrtrmlagssvm--scvscfflsfngsesamiallclfggvsia-swnald 670
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 446 VYTPDEVVPTATRALGLTGTCGSMARVGALI-----TPFIAQVMLESSVYLTAVYSGCCLL 500
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 vltvlypsdkrttafgfnalcklaavlgisftsvgitkaapilfasaalalgssia 730
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 501 AALASCFLPETKGGGIQ 518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 731 lklp-----etrgqvlq 742
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
AA92902
ID AAY92902 standard; Protein; 542 AA.
AC AAY92902;
XX
XX
XX 26-SEP-2000 (first entry)
XX Human cerebral organic anion transporter OAT3 protein.
XX Human; organic anion transporter protein; OAT3; cerebral tissue.
XX Homo sapiens.
XX
XX PN WO200017237-A1.
XX
XX PD 30-MAR-2000.
XX
XX PF 20-SEP-1999; 99WO-JP05120.
XX
XX PR 18-SEP-1998; 98JP-0265126.
XX
XX PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
XX PI Endou H, Sekine T, Kusuhsara H;
XX
XX DR WPI: 2000-283546/24.
XX
XX DR N-PSDB: AAL11146.
XX
XX PT Organic anion transporter protein OAT3 expressed in cerebral tissue,
XX antibodies to it and gene encoding it for study of cerebral transport
XX of anions including drug molecules
XX
XX PS Claim 2; Page 18-20; 48pp; Japanese.
XX
XX CC This sequence represents a human organic anion transporter protein OAT3.
XX The sequence is used to regulate the transport of anions in cerebral
XX tissue, including drug molecules, and regulation of this transport.
XX
XX SQ Sequence 542 AA;

Query Match 14.0%; Score 398; DB 21; Length 542;
Best Local Similarity 23.8%; Pred. No. 3.2e-35;
Matches 130; Conservative 82; Mismatches 190; Indels 144; Gaps 17;

QY 72 DAVEAIG-FGRFQWKLVLTLGLAW--MADAMEMMILSLAPQLHCE-----WRLP- 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 eildrvsgmhqfihvaigilpilnmanhnlqiftaatpvhhrpphnastgpgwlp 64
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 119 -----SNQVA-----L 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 gpngkperclrfvhpnaslpndtqrampcldgvyvynstkdsvtwdlvcnsnkikem 124
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 125 LTSVVVFCMMSSSLWGNISDQYGRKTKLTSVLWTLTYGILSAFAPYSWILVLRGLVG 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 aqsifmagilaggivlgdlsdrfgrripiltcsyllaasgsgaafspftplymvfrfclg 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 185 FGIGGVPOS-VTLVAEFLPMKARAKILLIEVFVAICTVFEVVLAVFVMPSLG-----WR 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 185 fgisgitlsvlinvewprrmrainstalgycytfuq-----fiipglayaipqwr 236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 239 WLLILSAVPLLPAVLFCFWLPESARYDVLSGNOEKAIALTKRIAT-----ENGAHPPLGK 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 wqlgtvsiipffvflsswwtpeisirwlvlskgssealkilrrvavfngkkeeerislee 296
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 294 LIISROED-----RGKMRDLF-TPHFRWTTLLLWFIFSNAPSYGLVLTTE----- 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 lklngkeislakakytasdlfrimlrmfclslawfatgfyyslamveefgvnly 356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 341 ----LFOAGDVCGISSRKKAVEAKSLACEYLSEEDYMDLLWTLTSEFPCVLVTWIDR 396
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 ilqiifggvdv-----pakfittilsly 379
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 397 IGRKKTMALCFVIFSCSL-LLFICVG-RNVLTLLLFIARAFISGGFOAAAYVTPVPT 454
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 lgrhtqaaallaggaillaltfvpldltvrtvlavfkgclsssfscfitytselypt 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 455 ATRALGLTGTCGSMARVGALITPFIQVMLESSVYLTAVYSGCCILAAALSCFLPIETKG 514
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 440 virqlgmgsnltwtrvgsmvslv-kitgevqpfipniilygitallgssaaflp-etin 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 515 GGLQES 520
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 qplpet 503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
AA92903
ID AAY92903 standard; Protein; 536 AA.
XX
XX AC AAY92903;
XX
XX DT 26-SEP-2000 (first entry)
XX
XX DE Rat cerebral organic anion transporter OAT3 protein.
XX
XX KW Rat; organic anion transporter protein; OAT3; cerebral tissue.
XX
XX OS Rattus sp.
XX
XX PN WO200017237-A1.
XX
XX PD 30-MAR-2000.
XX
XX PF 20-SEP-1999; 99WO-JP05120.
XX
XX PR 18-SEP-1998; 98JP-0265126.
XX
XX PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
XX PI Endou H, Sekine T, Kusuhsara H;
XX
XX DR WPI: 2000-283546/24.
XX
XX DR N-PSDB: AAL11147.
XX
XX PT Organic anion transporter protein OAT3 expressed in cerebral tissue,
XX antibodies to it and gene encoding it for study of cerebral transport
XX of anions including drug molecules
XX
XX PS Claim 2; Page 45-46; 48pp; Japanese.
XX
XX CC This sequence represents a rat organic anion transporter protein OAT3.
XX The sequence is used to regulate the transport of anions in cerebral
XX tissue, including drug molecules, and regulation of this transport.
XX
XX SQ Sequence 536 AA;

Query Match 13.9%; Score 396.5; DB 21; Length 536;
Best Local Similarity 28.3%; Pred. No. 4.6e-35;
Matches 119; Conservative 79; Mismatches 173; Indels 49; Gaps 15;

```

Query Match 13.98; Score 395; DB 18; Length 537;

[illegible][illegible]

RESULT 12

**6. Murine organic anion transporter 6.**

X N W0200070048-A1

R 12-MAY-2000; 2000US-0570293.



Db 380 lgrhrtqaalllaggallaltftvpidlqtvtvavfkgkclssstscilflytselypt 439  
Qy 455 ATRALGLTCSGMARVYCALITPTIAQVLMLESSVYLTAVYSGCLLAALASCFLEPIETKG 514  
Db 440 vtrgtmgvsnlwtvgvsmvplv-kitgevqfipnilygitaliggsaalilp etln 497  
Qy 515 GGLOES 520  
Db 498 qplpet 503  
RESULT 14  
AAB12131  
ID AAB12131 standard; Protein: 607 AA.  
AC AAB12131;  
XX 02-FEB-2001 (first entry)  
XX Hydrophobic domain protein from clone HP03092 isolated from Liver cells.  
XX Human; secreted protein; membrane protein; hydrophobic domain;  
XX proliferation control; differentiation induction; material transport;  
XX biophylaxis; signal receptor; ion channel; transporter; immunostimulant;  
XX immunosuppressant; haematopoiesis regulator; chemotactic; chemokinetic;  
XX haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;  
XX autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer.  
OS Homo sapiens.  
XX WO200029448-A2.  
XX 25-MAY-2000.  
XX 17-NOV-1999; 99WO-JP06412.  
XX 17-NOV-1998; 98JP-0326255.  
XX 22-DEC-1998; 98JP-0364315.  
XX 16-MAR-1999; 99JP-0069811.  
XX 27-APR-1999; 99JP-0119299.  
XX 19-MAY-1999; 99JP-0138169.  
XX (SAGA ) SAGAMI CHEM RES CENT.  
XX (PROT-) PROTEGENE INC.  
XX Kato S, Kimura T;  
XX WPI; 2000-387753/33.  
XX N-ESDB; AAA62004, AAA62014.  
XX proteins comprising hydrophobic regions, such as secretory and membrane  
XX proteins, useful in research and diagnostics and having various  
XX activities e.g. immunomodulatory, antiinflammatory, chemokinetic,  
XX hemostatic, thrombolytic -  
XX Claim 1; Page 236-238; 410pp; English.  
XX Secretory proteins play important roles in the proliferation control, the  
XX differentiation induction, the material transport and the biophylaxis of  
XX cells. Membrane proteins have important roles as signal receptors, ion  
XX channels and transporters. The present sequence is a human protein which  
XX has at least one hydrophobic domain. This protein may be a secretory or a  
XX membrane protein. The present protein may have cytokine and cell  
XX proliferation/differentiation activity, immune stimulating or suppressing  
XX activity, haematopoiesis activity, tissue growth activity,  
XX activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
XX and thrombolytic activity, anti-inflammatory activity and tumour  
XX inhibition activity. The present protein could therefore be used for  
XX treatment of autoimmune disease, Alzheimer's disease, Parkinson's  
XX disease, and cancer.  
XX sequence 607 AA;  
SQ

Query Match 13.3%; Score 379; DB 21; Length 607;  
Best Local Similarity 26.5%; Pred. NO. 5e-33;  
Matches 140; Conservative 76; Mismatches 204; Indels 108; Gaps 18;  
Qy 29 TASGEHEVQIEGVHVGLEAVELDD-----CAAVPKFANPTDDTFVWEVDAVEAIGFGKEQ 83  
Db 127 tiatesqvgiyihlevecrwrspweaagrglpweea-----eaaglgdrk 173  
Qy 84 -----WKLS---VLTGLAMMADAMEMMILSILAPOLHCEWRPLSQVAL---LTSVVVF 131  
Db 174 vsyspswreslgglisgmew-----dlvce-----qklnraastffa 212  
Qy 132 GMMSSSTLWGNISDOYGRKTKLISVLWTLTYGILSAFAPVYSWILVLRGLVGFIGGVP 191  
Db 213 gvlvgavafgylsdrfgrtrlllvayvstlvlglaasvvyvmfalftrltgsalagft 272  
Qy 192 QSV-TLYAEFLPMKARAKCILLIEVFWAIGTVEFVLAIVFVMPSLGMRWLLILLSAVPILL 250  
Db 273 ilmplelelvdehr:vagvlsstftwtggmllaivgyllr---dwrwillavtlpca 329  
Qy 251 FAVLCFMPLESARYDVLSGNOEKAIATLKRIATENGAP-----MPLGKLIIS 297  
Db 330 gllslwvwpesarwlltqghvkeahryllhcarlngprvcedsfseqeavskvaagerivr 389  
Qy 298 QEDRCMKMDLF-TPHFRWTLILLWFIFSNFASYGLVLLTTELFOAGDVCGISSRKKA 356  
Db 390 rp-----syldfrtrlrhslccvvvfwgvnfsyyqlsl-----dvsg----- 429  
Qy 357 VEAKSLACEYLSSEEDYMDLLTTLSEFPGLVLTWIIDRLGRKKTMA--LCFVIFSFC 414  
Db 430 -----lglnvyqtqlfgavelpskllvylsvryagrltqagtlgtalagft 478  
Qy 415 LLLPICVGRNVLTLLLFIAFAFISGGFOAAVYVTPETATRALGLTCSGMARVYCALI 474  
Db 479 rllvssdmkswstlavmgkafseafitayiftselvptvtrtgmglcalvgrlgs 538  
Qy 475 TPFTAQVWLESSVYVLT--AVYSGCCLLAALASCFLEPIETKGGLOES 520  
Db 539 aplaa-----lldgwslspkityggiallaagtalilp-etrqqlpet 582  
RESULT 15  
AAY44278  
ID AAY44278 standard; Protein: 550 AA.  
XX AAY44278;  
XX 29-FEB-2000 (first entry)  
XX Human organic anion transporter.  
XX Human organic anion transporter; hOAT; nephrotoxic compound; screen;  
XX drug-drug interaction; nucleotide phosphate.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX Peptide 515..528  
XX /label= Immunogen  
XX /note= "For producing antibodies"  
XX WO9964459-A2.  
XX 16-DEC-1999.  
XX 10-JUN-1999; 99WO-US13172.  
XX 11-JUN-1998; 98US-0088864.  
XX 03-MAY-1999; 99US-0132267.  
XX (GILE-) GILEAD SCI INC.  
XX

Cihlar T:

XX WP1: 2000-097519/08.  
DR N-PSDB; AAZ29300.

XX Human organic anion transporter nucleic acid used in drug screening,  
PT identifying drug-drug interactions or individuals in whom particular  
PT drugs are nephrotoxic

XX Claim 9; Fig 1: 36pp; English.

XX The present sequence is human organic anion transporter (HOAT) protein.  
CC HOAT is expressed in kidney and brain and removes toxic anions  
CC from the circulation. It can be produced in host cells by transforming  
CC them with recombinant vectors containing nucleic acid encoding HOAT.  
CC It can be used in screening for compounds that suppress or enhance anion  
CC uptake and transport by HOAT, to identify molecular variants of  
CC nephrotoxic compounds, to detect drug-drug interaction within kidney  
CC and brain and to test for transport of a covalently modified form of a  
CC nucleotide phosphonate analogue by HOAT.

XX Sequence     550 AA;

Query Match                 13.2%; Score 374.5; DB 21; Length 550;  
Best Local Similarity      25.6%; Pred. No. 1.4e-32;  
Matches 132; Conservative 86; Mismatches 207; Indels 91; Gaps 18;

QY 54 AAVPKETANPIDDFFMWEDA-----VEAIGTCKFOWKLSVLTGL----- 92  
||| : | : : : : ||| : |  
Db 42 aaaihrrcpadaniskngglewlprdrqgqpescrlftspwgipfingteangta 101  
  
QY 93 -----AWMDAMEMMILSLAPOLHCEWRLPMSQWA---LLTSVFVGMMSSSTLWGN1 143  
||| : | : : : : ||| : | : : : :  
Db 102 tepctdgiydn-----stfpstvltewdlvcsrhralqlaqslgmvgllgamvfyl 155  
  
QY 144 SDQGRKTGKLVLWLTVLYGISAFAPVYSNIIIVRLGVFGGGVP-QSVTLIAEFLP 202  
:||| : ||| : | : |||| : ||| : | : | : | : ||| : ||| :  
Db 156 adlgrkrkvilnylqtavsgtcaafapnfyicafllgmalagisinmctinvevp 215  
  
QY 203 MKARAKILLLEFPWAJTFEVVVLAHVMPISLGWRMLLIISAYPLLLFAVLCEFWLPESA 262  
||| : ||| : : ||| : ||| : | : ||| : | : | : | : ||| :  
Db 216 iht racvtgllygyyslqq-fllagvayaph--wrhlqlivsapffaffiywffiesa 272  
  
QY 263 RYDLVSGNQEKAIATLKRIA-----TNGAMPPLGKKLIISRQB--RKG-----MRDIFT 310  
| : ||| : | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
Db 273 rwhssgtldltralgrvaringkreegaklsmeviraslkeltmgkgqasameilir 332  
  
QY 311 PHFRWTLLLWFIFSNASFYYGIUVLLTTTELFQAGDVCGISSRKKAWEAKCSLACEYLSE 370  
| : | : | : | : ||| : ||| : | : | : | : ||| : ||| :  
Db 333 ptlrhlfclsmfwatsfayglm-----dqggf-----vsyiqliq 371  
  
QY 371 EDY--MDLLTWLTSSEPGVGLVTLMIDLRGKRKMALCFVITSPCSSLLLFCVGRN--VL 426  
: || : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
Db 372 vifgavl-----paklvfgvlinslrgprpqmaalllaglicillingvipqdqsivr 423  
  
QY 427 LLLLLIARAFTSFGFOAAVYVTPPEVYTATRALLGTCTSGMARVCALITPIPAQVMLESS 486  
| : | : | : | : ||| : ||| : | : | : | : ||| : ||| :  
Db 424 tslavikgcilaasnfcilytgelytmirqtgmgtmarvygsvisplvs----mcae 480  
  
QY 487 VY--LTLAVYSGCCLLAALASCFPIETKGGLOES 520  
: | : ||| : | : ||| : | : ||| : | : ||| : | : ||| :  
Db 481 lypsmplfiygavpaavasavtlip-etligqlpdtt 515



Wed Mar 13 13:08:08 2002

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:40:14 ; Search time 21.65 seconds  
(without alignments)  
569.598 Million cell updates/sec

Title: US-09-911-667A-2  
Perfect score: 2846  
Sequence: 1 NEEDLFQLRLPVVKERRTG.....MVGRGMHGAGVTRNSGSQE 548

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	429	15.1	555	3 US-08-501-572-3	Sequence 3, Appli
2	429	15.1	555	3 US-09-040-444-3	Sequence 3, Appli
3	426.5	15.0	556	3 US-08-501-572-1	Sequence 1, Appli
4	426.5	15.0	556	3 US-09-040-444-1	Sequence 1, Appli
5	425.5	15.0	553	3 US-08-501-572-2	Sequence 2, Appli
6	425.5	15.0	553	3 US-09-040-444-2	Sequence 2, Appli
7	395	13.9	537	2 US-08-647-397-2	Sequence 2, Appli
8	217	7.6	584	2 US-08-928-692-13	Sequence 13, Appli
9	203.5	7.2	520	4 US-08-968-127-2	Sequence 2, Appli
10	203.5	7.2	524	2 US-08-928-692-12	Sequence 12, Appli
11	178	6.3	493	2 US-09-031-392-10	Sequence 10, Appli
12	178	6.3	493	4 US-09-299-549-10	Sequence 10, Appli
13	171	6.0	494	2 US-09-031-392-5	Sequence 5, Appli
14	171	6.0	494	4 US-09-299-549-5	Sequence 5, Appli
15	157	5.5	488	2 US-08-928-692-11	Sequence 11, Appli
16	154.5	5.4	500	2 US-09-031-392-7	Sequence 7, Appli
17	154.5	5.4	500	4 US-09-299-549-7	Sequence 7, Appli
18	154.5	5.4	509	2 US-09-031-392-6	Sequence 6, Appli
19	154.5	5.4	509	4 US-09-299-549-6	Sequence 6, Appli
20	150.5	5.3	534	2 US-09-031-392-4	Sequence 4, Appli
21	150.5	5.3	534	4 US-09-299-549-4	Sequence 4, Appli
22	147.5	5.2	492	2 US-08-355-844-3	Sequence 3, Appli
23	147.5	5.2	492	5 PCT-US95-16126-3	Sequence 3, Appli
24	147	5.2	488	2 US-08-928-692-10	Sequence 10, Appli
25	145	5.1	563	2 US-09-031-392-2	Sequence 2, Appli
26	145	5.1	563	4 US-09-299-549-2	Sequence 2, Appli
27	138	4.8	423	2 US-08-494-907-14	Sequence 14, Appli

28	138	4.8	423	5 PCT-US96-10986-14	Sequence 14, Appli
29	135.5	4.8	383	2 US-09-031-392-3	Sequence 3, Appli
30	135.5	4.8	383	4 US-09-299-549-3	Sequence 3, Appli
31	133	4.7	322	4 US-08-964-127-6	Sequence 6, Appli
32	131.5	4.6	457	2 US-08-882-704A-6	Sequence 6, Appli
33	124	4.4	286	4 US-08-964-127-4	Sequence 4, Appli
34	116	4.1	3033	1 US-07-925-695-5	Sequence 5, Appli
35	112	3.9	109	2 US-08-647-397-4	Sequence 4, Appli
36	112	3.9	455	1 US-08-035-928-2	Sequence 2, Appli
37	111.5	3.9	1820	3 US-07-998-289B-8	Sequence 8, Appli
38	111.5	3.9	2100	2 US-08-808-793-23	Sequence 23, Appli
39	111.5	3.9	2100	3 US-08-772-512A-19	Sequence 19, Appli
40	109.5	3.8	418	4 US-09-030-267-5	Sequence 5, Appli
41	108	3.8	1976	3 US-09-024-020B-9	Sequence 9, Appli
42	108	3.8	1978	3 US-09-024-020B-3	Sequence 3, Appli
43	108	3.8	1988	3 US-09-024-020B-4	Sequence 4, Appli
44	106	3.7	569	2 US-08-750-723A-2	Sequence 2, Appli
45	106	3.7	569	4 US-09-191-275-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-501-572-3  
; Sequence 3, Application US/08501572  
; Patent No. 6063623  
; GENERAL INFORMATION:  
; APPLICANT: Koepsell, Hermann  
; APPLICANT: Gorboulev, Valentin  
; TITLE OF INVENTION: Transport Protein Which Effects The  
; TITLE OF INVENTION: Transport Of Cationic Xenobiotics and/or Pharmaceuticals,  
; TITLE OF INVENTION: DNA Sequences Encoding It And Their Use.  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan,Henderson,Farabow,Garrett & Dunner  
; STREET: 1300 I Street, N.W., Suite 700  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3315  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/501,572  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Toohey, Kimberlin M  
; REGISTRATION NUMBER: 35,391  
; REFERENCE/DOCKET NUMBER: 02481.1453-00000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)408-4000  
; TELEFAX: (202)408-4400  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 555 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-501-572-3

Query Match 15.1%; Score 429; DB 3; Length 555;  
Best Local Similarity 29.4%; Pred. No. 3.6e-39;  
Matches 126; Conservative 70; Mismatches 157; Indels 76; Gaps 14;  
QY 119 SWQVALITSVVFGVMSSSTLWGNISDQYGRKTKGLKISVLWTLTYGILSAPVYSWTLV 178

```

; MOLECULE TYPE: peptide
; US-09-040-444-3

Query Match 15.1% Score 429; DB 3; Length 555;
Best Local Similarity 29.4%; Pred. No. 3.6e-39;
Matches 126; Conservative 70; Mismatches 157; Indels 76; Gaps 14;

QY 119 SNOVALLTSVVFVGMSSSTLWGNISDOYGRKTKLGKISVLMTLYYGIILSAFAPVYSWLV 178
DB 146 SMLDLFOSSVNVGFFTCGSMISGYIADRFGRKCLLTIVLINAAGVLMALISPTITWMLI 205
QY 179 LR---GLV---GFGIGGVPOSVTLYAEFLPMKAKAKCILLIEVFWAIGTVFVVLV---VF 230
DB 206 FRLIOGLVSKAGWLIIGYI-----LITEVGGRYRRTVGIYQVAYTVGL---LVLAGVAY 257
QY 231 VMSLGRWMLILSAVPLLLFAVLCFWLPESARYDVLGNOEKATATLKRIATENGAPMP 290
DB 258 ALPH--WNLQFTVALPNFFLLYYWCIPESPRWLIISONKNAEAMRIIKHIAKNGKSLP 315
QY 291 LCKLIISROEDRGK-----MRDLF--TPHFRWTTLLLMFIFSNFASYGLVLLTTELFQA 344
DB 316 ASLOQLRLLEETGKKNLPSFLDLVTRTPQIRKHTMILMYNMTSSVLYOGLIM---HMGLA 372
QY 345 GDVCGISSRKKRAVEAKCSLACEYSEEDYMDLLMTTISEFFPGVLTWIIDRLGRKKTMA 404
DB 373 GD-----NIYLDFFYSALVEFFPAFMIIILIDRIGRRYPWA 408
QY 405 LC-FVIFSCSLLFI-----CVGRNVLTLLLFIAFAFISGFGFOAAVYVTPEV 451
DB 409 ASNMVAGAACLASVPIPGDLQWLKIIISCLGRMGITM-----AYEIVCLVNAEL 457
QY 452 YPTATRALGLGTCSGMARVAGALITPFIQVLMLESSVYLTAVYSGCCLLAALASCFLPIE 511
DB 458 YPTFIRNLGVHICSSMCDIGGIITPFLVYRLTNWLELPLMVFGVLGVAGGLALLLP-E 516
QY 512 TKGGGLQES 520
DB 517 TKGRALPET 525

RESULT 3
US-08-501-572-1
; Sequence 1, Application US/08501572
; Patent No. 6063623
; GENERAL INFORMATION:
; APPLICANT: Koepsell, Hermann
; APPLICANT: Grundeman, Dirk
; APPLICANT: Gorboulev, Valentin
; TITLE OF INVENTION: Transport protein which effects the
; TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,
; TITLE OF INVENTION: DNA sequences encoding it and their use.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: March 18, 1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: O'Connor, Steven P
; REGISTRATION NUMBER: 41,225
; REFERENCE/DOCKET NUMBER: 2481.1453-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)408-4000
; TELEFAX: (202)408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 555 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; REFERENCE/DOCKET NUMBER: 02481.1453-00000
```



STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/501,572  
FILING DATE: 2000-05-31  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Toohy, Kimberlin M  
REGISTRATION NUMBER: 35,391  
REFERENCE/DOCKET NUMBER: 02481.1453-00000  
TELEPHONE: (202)408-4000  
TELEFAX: (202)408-4400  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-501-572-2

Query Match 15.0%; Score 425.5; DB 3; Length 553;  
Best Local Similarity 30.2%; Pred. No. 8.7e-39;  
Matches 127; Conservative 66; Mismatches 167; Indels 61; Gaps 11;

Qy 119 SQQVALLTSVVFVGMSSSTLWGNISDOYGRKTGLKISVLTLYGILSAFAPVYSWILV 178  
Db 145 SSKLDLFQSCLNAGFFGSLGVGYFADRGKRLCLLGTVLVNAVSGVLMAFSPNYMSMLL 204  
Qy 179 LRGLVGF-GIGGVPOSVTLYAEFLPMKARAKCILLIEVFMAIGTVFEVVLAVFVMPISLW 237  
Db 205 FRLLQGLVSKGNMAGYTLITEFVSGSRRTVAIMYQMAFTVGLVALTGLA-YALPH--W 261  
Qy 238 RWLLILSAVPLLLFAVLCFWLPESARYDVLGNOEKAIAITLKRIATENGAPMLGLKLIIS 297  
Db 262 RWLQVAVSLPTFLFLYYWCVPSPRWLLSQKRNTAEIKIMDHIAOKNGKLPADLKMLS 321  
Qy 298 RQED-----RGKMRDLF-TPHFRWTTLLWFIWFSNAFSGYGLVLTTLTFLQAGDVCGIS 351  
Db 322 LEEDVTEKLSFADLPTPRLRKRRTFILMYLWFTDSVLYQGLIL-----HMGATSG-- 373  
Qy 352 SRKKAWEAKCSLACEYLSEEDYMDLLWTTLSFPPGVLTWIIIDRLGRKKTMAI----- 405  
Db 374 -----NLYDLFLYSALVEIPGAFIALITIDRVGRYIPMAVSNLLAG 414  
Qy 406 --CFVIFSFCSL-----LLFICVGRNVLTLLFIARAFISGGFQAAYVYTPVPTATRAL 459  
Db 415 AACLVIFISPDHLWLNIIIMCVGRMGITAI-----OMICLVNAELYPFTVRNL 463  
Qy 460 GLGTCGSMARVAGALITPFIQAQVMLESSVYLT LAVYSGCCLLAALASCFLPIETKGGLOE 519  
Db 464 RMVVCSSICDIGGIITPFIIVRREVWQALPLILFAVLGLLAAGVTLLLP-ETKGDALPE 522  
Qy 520 S 520  
Db 523 T 523

RESULT 6  
US-09-040-444-2  
; Sequence 2, Application US/0904044  
; Patent No. 6063766  
; GENERAL INFORMATION:

APPLICANT: Koepsell, Hermann  
APPLICANT: Grundeman, Dirk  
APPLICANT: Gorboulev, Valentin  
TITLE OF INVENTION: Transport protein which effects the  
TITLE OF INVENTION: Transport of cationic xenobiotics and/or pharmaceuticals,  
TITLE OF INVENTION: DNA sequences encoding it and their use.  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P.  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,444  
FILING DATE: March 18, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: O'Connor, Steven P  
REGISTRATION NUMBER: 41,225  
REFERENCE/DOCKET NUMBER: 2481.1453-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)408-4000  
TELEFAX: (202)408-4400  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-040-444-2

Query Match 15.0%; Score 425.5; DB 3; Length 553;  
Best Local Similarity 30.2%; Pred. No. 8.7e-39;  
Matches 127; Conservative 66; Mismatches 167; Indels 61; Gaps 11;

Qy 119 SQQVALLTSVVFVGMSSSTLWGNISDOYGRKTGLKISVLTLYGILSAFAPVYSWILV 178  
Db 145 SSKLDLFQSCLNAGFFGSLGVGYFADRGKRLCLLGTVLVNAVSGVLMAFSPNYMSMLL 204  
Qy 179 LRGLVGF-GIGGVPOSVTLYAEFLPMKARAKCILLIEVFMAIGTVFEVVLAVFVMPISLW 237  
Db 205 FRLLQGLVSKGNMAGYTLITEFVSGSRRTVAIMYQMAFTVGLVALTGLA-YALPH--W 261  
Qy 238 RWLLILSAVPLLLFAVLCFWLPESARYDVLGNOEKAIAITLKRIATENGAPMLGLKLIIS 297  
Db 262 RWLQVAVSLPTFLFLYYWCVPSPRWLLSQKRNTAEIKIMDHIAOKNGKLPADLKMLS 321  
Qy 298 RQED-----RGKMRDLF-TPHFRWTTLLWFIWFSNAFSGYGLVLTTLTFLQAGDVCGIS 351  
Db 322 LEEDVTEKLSFADLPTPRLRKRRTFILMYLWFTDSVLYQGLIL-----HMGATSG-- 373  
Qy 352 SRKKAWEAKCSLACEYLSEEDYMDLLWTTLSFPPGVLTWIIIDRLGRKKTMAI----- 405  
Db 374 -----NLYDLFLYSALVEIPGAFIALITIDRVGRYIPMAVSNLLAG 414  
Qy 406 --CFVIFSFCSL-----LLFICVGRNVLTLLFIARAFISGGFQAAYVYTPVPTATRAL 459  
Db 415 AACLVIFISPDHLWLNIIIMCVGRMGITAI-----OMICLVNAELYPFTVRNL 463  
Qy 460 GLGTCGSMARVAGALITPFIQAQVMLESSVYLT LAVYSGCCLLAALASCFLPIETKGGLOE 519  
Db 464 RMVVCSSICDIGGIITPFIIVRREVWQALPLILFAVLGLLAAGVTLLLP-ETKGDALPE 522  
Qy 520 S 520  
Db 523 T 523

Db 523 T 523

QY 133 MMSSTLWGNISDOYGRKTKGLKISVLWTLTYGILSAFAPVYSWILVLRGLVGFIC-GVP 191  
DB 139 ALITISFAGTAADIFGRKCLMGSLNMFVIGAILQVSAHTFWMAGVRLIMFGVGIGSL 198  
QY 192 QSVTLVAEFLPMKARAKCILLIEVEFWAIGTFEVVL-----AVFVPSLGRWLLIISAVP 247  
DB 199 IAPLFISEIAPKMKR-LTVINSLWLTGG--QLVAYGGAGLNVNNGRILVGLSLIP 255  
QY 248 LLL-FAVLFWLPSARYDVLGNSQKATATLKRIATENGAPM---PGLKLIISROEDRG 303  
DB 256 TAVQFTCLCF-LPDTPRYVMKGLARATEVLKRSYTDTSSEIIERKVFELVTLNQSIPG 314  
QY 304 KMRDLTPHFRWTLTLWFIFWESNAFYSGLVLLTTELFQAGDVCGISRRKKAVERKCSL 363  
DB 315 KN-----VPEKVNWTI-----KELHTVPSNLRALIIGCGL 344  
QY 364 -ACEYLSEEDYMDLLWTLTSEPPG-----VLVTLWIIDRL 397  
DB 345 QAIQOFTG-----WNSLMYFSGTIFETVGRKNSSAVSIIVSGTNFIETLVAFFSIDKI 397  
QY 398 GRK-----KTMAL--CFVIFSPC-----SLLLFICVGRN-----VLTLLLFIFARAF 436  
DB 398 GRRTILLIGPGMTALVWCSTAFHGLKFGCAVAVVSSGFWIIVFIIVFAAF 457  
QY 437 ISGFOAAYVYTPVYPTATRALGLGTCGSMARVGAITPFAQVMSLSVYL-TLAVYS 495  
DB 458 YALGIGTVPMQOSELFPQNVRGIGTSYATATNAGSLVIASITFLTMLQNTITPAGTFAPFA 517  
QY 496 GCCLLAALASCFLPIETKGGGLQE 519  
DB 518 GLCLSTIFCYFCYPELSGLEEE 541

RESULT 9

US-08-964-127-2  
; Sequence 2, Application US/08964127  
; Patent No. 6277565  
; GENERAL INFORMATION:  
; APPLICANT: Grandearl, Andrew David John  
; TITLE OF INVENTION: NOVEL GENES ENCODING TRANSPORTER-LIKE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/964,127  
; FILING DATE: 06-NOV-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crews, Ph.D., L. Lee  
; REGISTRATION NUMBER: P-43,567  
; REFERENCE/DOCKET NUMBER: 07334/038001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 520 amino acids  
; TYPE: amino acid

; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-964-127-2

Query Match 7.2%; Score 203.5; DB 4; Length 520;  
Best Local Similarity 25.6%; Pred. No. 5.7e-14;  
Matches 116; Conservative 68; Mismatches 174; Indels 95; Gaps 24;

QY 104 LSILAPQLHCEWRLP---SWQVALLTSVYFVCGMSSSTLWGNISDOYGRKTKGLKISVLWT 160  
DB 79 LPVLTNAIGOWDLVCDLGMQVLEQILFILGFASGYFLGYPADRFGR--GI---VLIT 134  
QY 161 LVY-----GILSAFAPVYSWILVLRGLVGFICGVPQSVTLV-AEFLPMKARAKCILLIEV 215  
DB 135 LGLVPCGVGGAAAGSSTGVMALELIGLFLAGVDLGVYLMRLLELCPDQRLKVALAGEL 194  
QY 216 FWAICTVFEVLAVFVMPSLGRWLLIISAVPLLIIFAVLCFWLP-----ESARYDVLGSGNQ 271  
DB 195 ---VGVGHFLEFLGALVSKDMRFLOQMITAPCILF--LFYGMWPGFLFESARMLIVKROI 249  
QY 272 EKAIATLKRIATENGAPMPLGKLIISROEDRGCKMRDL-----FTPHFRWTLTLL---L 320  
DB 250 EEAQSVLRILABRN---RPHGQML--GEEAQALQDLENTCPLPATSSFSFASILLNYRNI 304  
QY 321 W-----FIFMSNAFYSYGLVLLTTELFQAGD-----VCGISSRKKKAVEAKCSLACEYLSSE 371  
DB 305 WKNLLILGFTN-FIAHAIRHCYQPVGGGSPSDFYLCSLASGTA-----ALACVFL--- 355  
QY 372 DYMDLLWTLTSEPPGVVLTWIIDRLGRK.MALCFVIFSPCSLLLLFICVGRNVLTLTLIF 431  
DB 356 -----GVTV-----DRFRRGILLISMTLTGIASLVL-----LGLWDY 388  
QY 432 IARAFIS-----GGF--QAAYVYT-----PEVYPTATRALGLGTCGSMARVGAITPFAIAQ 480  
DB 389 LINEAALTFSVLGLFSSQAAAILSTLLAAEVIPTVYRGRGLGLMALGALGGLSGP-AQR 447  
QY 481 VMLESVYLTAVYSGCCLLAALASCFLPIETK 513  
DB 448 LHMHGAFLOHVVLAAACALLCILSIMLLP-ETK 479

RESULT 10

US-08-928-692-12  
; Sequence 12, Application US/08928692  
; Patent No. 5958727  
; GENERAL INFORMATION:  
; APPLICANT: Brody, Howard  
; APPLICANT: Yaver, Deborah S.  
; APPLICANT: Lamsa, Michael  
; APPLICANT: Hansen, Kim  
; TITLE OF INVENTION: Methods for Modifying the Production of  
; TITLE OF INVENTION: a Polypeptide  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5958727o No. 5958727disk of No. 5958727th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928,692  
; FILING DATE: 12-SEPT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J

```
;
;
;   REGISTRATION NUMBER: 33,728
;   REFERENCE/DOCKET NUMBER: 4944,200-US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-867-0123
;   TELEFAX: 212-878-9655
;   INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 524 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: No. 5958727e
;
US-08-928-692-12

Query Match      7.2%; Score 203.5; DB 2; Length 524;
Best Local Similarity 22.0%; Pred. No. 5.7e-14;
Matches 119; Conservative 81; Mismatches 197; Indels 143; Gaps 22;

QY 48 VELDDGAAPKFEFANPTDD-----TFMVEDAVEAIGFGKQKLSVLTGLAW 94
DB 48 VPLDDRKAANNVYINSDDELPTTISYNNPKPTPWAEEETVAAAALITMLWSLV----- 101
QY 95 MADAMEMMILSIAPLQHCWRPLSNQVALLTSVVFVGMSSSTLMGNISDOYGRKTGLK 154
DB 102 -----SSFAVGGMGTASFFGGWGLDGLTGRKAML 129
QY 155 ISVLWTLYYGILSAFA--PVYSWILVLRGLVGGVGPVOSTVLY-AEFLPMKARAKCI 210
DB 130 VANLISLVGALLMGKSGKLGPSHILITAGRSISGLYGLISGLVPMYIGEIAPTALRG-- 186
QY 211 LLIEVFMAIGTFVFEVLIAPVMP-----LG-----WRMLLISAVPLLLFAVLFCWL 258
DB 187 -----ALGTFHQLAIVTGLITISIIQITGLEFILCNVDLHILLGLSGVAILQSLLFFC 239
QY 259 PESARYDVLSONQE-KAIAATKRIATENGAPMLGLKLIISROEDRGKMR-----DLFT-PH 312
DB 240 PESPRYLKILDEEVKAKSLRUGYDDVTKDINEMKEREASSEQKVSIIQLFTNSS 299
QY 313 FRWTTLLWFTWFSNAPS-YVGLVLLTTELFQAGDVGCGISSRKKAVEAKCSLACEYLSEE 371
DB 300 YROPILVALMLHVAQFGSGINGIFYSTSIFQ---TAGIS---KPVYATIGVGA----- 347
QY 372 DYMDLLTTLTSEFPGLVLTWIDRLGRKKTMAICLVIFSCSLLLFICVGR----- 423
DB 348 ---VNVFTAVSVF-----LVEKAGRRSLFLIGMSGMFVCA--IPMSGVLVLLNKFSW 395
QY 424 ----NVLTLTLLFIARFTSGG----FOAAVYVT--PEVYPTATRALGLGTCGSMARVYAL 473
DB 396 MSYVSMIAIFLVFSFFELGPGPIPMWVAEFFSGPRPAALATAAESNWTCTNFIV---AL 452
QY 474 ITPFIAQWMLSESVYLVAVYSGCCLLAALASCFLPIETKGGGLQE-----SSHR 523
DB 453 CFQYIADFCGPVVFEL-----FAGVLLAFTLTFFKVPETKGSFEETAABEFQKSGSAHR 508

RESULT 11
US-09-031-392-10
; Sequence 10, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
```

```
;
;
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: Windows95
;   SOFTWARE: FastSeq for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/031,392
;   FILING DATE: 26-FEB-1998
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Meiklejohn, Ph.D., Anita L.
;   REGISTRATION NUMBER: 35,283
;   REFERENCE/DOCKET NUMBER: 07334/072001
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617/542-5070
;   TELEFAX: 617/542-8906
;   TELEX: 200154
;   INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 493 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-09-031-392-10

Query Match      6.3%; Score 178; DB 2; Length 493;
Best Local Similarity 21.7%; Pred. No. 3.7e-11;
Matches 110; Conservative 92; Mismatches 196; Indels 108; Gaps 23;

QY 67 TFMVEDAVEAIGFGKQKLSVLTGLAWMADAMEMMILSIAPLQHCWRPLS----- 119
DB 13 TLVLAVLIAALGSGFOXYNLGVI-----NAPQKVEAFYETWLGXGXSPVTLTL 65
QY 120 MQVALLTSVVFVGMSSSTLMGNISDOYGRKTGLKLSVLTWLYYGLISAFAPV---YSWI 176
DB 66 W--SLSVISFAGVGMGTSGFLVCGXIGNLRGKXAMLVNNVNLATAGGLMLGLAKXAXSEF 123
QY 177 LVLRGLVGGVGGVGPVOSTVLY-AEFLPMKARAKCILLIEVFMAIGTFVE--VVLA 233
DB 124 ILGRFTIIGLYCGLSSGVPMYVGEISPTALRG-----ALGTNLQGLVIGILIAQ 173
QY 234 SLG-----WRMLLISAVPLLLFAVLFCWLPESARYDVLSONQE-KAIAATLKRIA 282
DB 174 VLGLDLSLGNESLWPLLLGLTGVPAQLLQLLLPFCPESPRYLLINKNEAKAKKALQRL 233
QY 283 TENGAMPPLGKLIISROEDRG-----KMRDLF-TPHRWTTLLWFIWFS-----NA 328
DB 234 GTADVSOEVAEM---KDESXMXSEKXSVLELFRSRXYRQPVIIVLQSLQSLSCINA 290
QY 329 FSYVGLVLTTELFQAGDVGCGISSRKKAVEAKCSLACEYLSEEDYMDLLTTLTSEFPGLV 388
DB 291 VFYV---STSIPEKAGV-----GQPVYATIGAG-----VNVTVTVSVF----- 327
QY 389 VTLMWIDRLGRKKTMAICLVIFSCSLLLFICVGR-----NVLTLTLLFIARAFIS 438
DB 328 ----VVERAGRRTLLHLLGLGMAGCAVLMTIALALLDQVPMMSVYSVIAIFGVFAFEVG 383
QY 439 GG-----FOAAVYVTPEVYPTATRALGLGTCGSMARVYALITPFIAQWMLSESVYLVAVY 494
DB 384 PGPIPMFIVAELEFSQGRPAIAVAGFSNNWTSNFIVG-LLFQYIAE-LLGPGYVFI 440
QY 495 SGCCLLAALASCFLPI-ETKGGGLQE 519
DB 441 ---LLLLFFITFLKVPETKGRITFDE 463

RESULT 12
US-09-299-549-10
; Sequence 10, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
```

;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
;; TITLE OF INVENTION: GLUTEX AND USES THEREOF  
;; NUMBER OF SEQUENCES: 10  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fish & Richardson P.C.  
;; STREET: 225 Franklin Street  
;; CITY: Boston  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02110-2804  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; OPERATING SYSTEM: Windows95  
;; SOFTWARE: FASTSEQ for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/299,549  
;; FILING DATE: 26-APR-1999  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 09/031,392  
;; FILING DATE: 26-FEB-1998  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Meiklejohn, Ph.D., Anita L.  
;; REGISTRATION NUMBER: 35,283  
;; REFERENCE/DOCKET NUMBER: 07334/072002  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617/542-5070  
;; TELEFAX: 617/542-8906  
;; TELEX: 200154  
;; INFORMATION FOR SEQ ID NO: 10:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 493 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-299-549-10

Query Match 6.3%; Score 178; DB 4; Length 493;  
Best Local Similarity 21.7%; Pred. No. 3.7e-11;  
Matches 110; Conservative 92; Mismatches 196; Indels 108; Gaps 23;  
QY 67 TMVEDAVEAIGFGKQKLSVLTGLANMADAMEMMILSILAPQLHCEWRFLPS-----119  
Db 13 TLVLAVLIAALSGFQYGNLGI-----NAPQKVEAFYETWLGKXGXPVPTLTL 65  
QY 120 WQVALLTSVVFVGMSSSTLWGNISDQYGRKTGLKISVLWTLTYGILSAFAPV---YSWI 176  
Db 66 W--SLVSIFAYGGMIGSFLVXIGNRLGRKXAMLVNNVLAIGAGLLMLGLAKXAXSFEM 123  
QY 177 LVLRLGLVFGIGVQPSVTLY-AEFLPMKARAKCILLIEVFWAIGTVFE--VVLAVFVMP 233  
Db 124 ILGRFIIGLYCGLSGVPMVYGEISPTALRG-----ALGTLNQLGIVIGILIAQ 173  
QY 234 SLG-----NRWLLTILSAVPLLLFAVLFCWLPESARYDVLGSGNE-KAIAATLKRIA 282  
Db 174 VLGLDSLLGNESLWPLLLGLTGPALLQLLLPFCPSRYLLINKNEARAKKALQRLR 233  
QY 283 TENGAPMPGLKLIISROEDRG-----KMRDLF-TPHFRWTLTLLWFIFWS-----NA 328  
Db 234 GTADVSQVEAEM---KDSRXMXSEKXSVLELFRSRYRQPVIIAIVLSQLSGINA 290  
QY 329 FSYGLVLLTTELFQAGDVCGISSRKKAVEAKCSLACEYLSIEDYMDLLTTLSEFPGVL 388  
Db 291 VFY-----STSEFKAGV-----GQPVYATIGAG-----VNVTVFTVVSF-----327  
QY 389 VTLWIIDRLGRKTKWALCFVTFSCSLLLFICVGR-----NVLTLILFIARAFIS 438  
Db 328 ----VVERAGRRTLHLGLGGMACCAVMTIALALDDQVPMMSYVSIVAIFGVFAFFVG 383  
QY 439 GG-----FOAAYVTVPEVYATRALGLGTCGSMARVALITPFAQVMESSVYLTAVY 494  
Db 384 PGPIPFIVAEILFSQGPRAATAVAGFSNWTNFIIVG-LLFOYIAE-LLGPVVFIVFAV- 440

QY 495 SGCCLLAALASCFLPI-ETKGGGLQE 519  
Db 441 ---LLLLFFIITFLKVPETKGRTFDE 463  
RESULT 13  
US-09-031-392-5  
; Sequence 5, Application US/09031392  
; Patent No. 5942398  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Weng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/031,392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 494 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-031-392-5

Query Match 6.0%; Score 171; DB 2; Length 494;  
Best Local Similarity 23.7%; Pred. No. 2.3e-10;  
Matches 123; Conservative 82; Mismatches 205; Indels 110; Gaps 27;  
QY 73 AVEAIGFKQKLSVLTGLANMADAMEMMILSILAPQLHCEWRFLPSWOVALLTS-----127  
Db 13 ASIAITIGSFQGYN--TG---VINAPEAIKDFLNTYLEERSETPPSV-LTSLWSLS 66  
QY 128 --VVFVGMSSSTLWGNISDQYGRKTGLKISVLWTLTYGILSAFAPVYSMILVL---RGL 182  
Db 67 VAIFSVMGIGSFGVLFVNRFGRRNSMLIVNLLAIGAGCLMGFCCKIAESVEMLILGLRI 126  
QY 183 VGFVGIGVQPSVTLY-AEFLPMKARAKCILLIEVFWAIGTVFE--VVLAVFVMPSLG---236  
Db 127 IGLFGLCTGTFVPMYIGESPTALRG-----AFGLNQLGIVIGILVAIFGLKV 176  
QY 237 -----NRWLLTILSAVPLLLFAVLFCWLPESARYDVLGSGNE-KAIAATLKRI-APENGA 287  
Db 177 ILGTEDLWPLLLGTFILPAIIQCAALPFCPSRPELLINKKEEKAKEILQRLWGTEDVA 236  
QY 288 ----PMPGLKLIISROEDRGKMRDLF-TPHFRWTLTLLWFIFWS-----NAFSYGLVL 336  
Db 237 QDIQEMKDESMMWS-QEKQVTVLELFRAPNRPQIISIMLQLSQOLSGINAVFY----291  
QY 337 LTTELFQAGDVCGISSRKKAVEAKCSLACEYLSIEDYMDLLTTLSEFPGVLVTLWIIDR 396



Db 292 -STGIFKADGV-----QEPVYATIGAG-----VVNTFTVSVF-----LVER 328  
QY 397 LGRKKT-----MALCFVIFSCSLL-----FICVGRNVLTLLLFIAFAFISGG- 440  
Db 329 AGRTLHLGLGMAFCSILMTI-SLLKDNYSWMSFICIG-----AIVFVAFVFEIGPGP 383  
QY 441 ---FOAAVYVTPVPTATRALGLGTCGSMARVAGALITPFIQAQVLMLESSVYLILAVYSGC 497  
Db 384 IPWFIVAEELFGQGPRAAMAVAGCSNWTSNFLVGLLFPSS--ATFYLGAYVFIVFTVF--- 438  
QY 498 CLLAALASCLPIETKGGGLQESSHREWGOMVG-RGMHG 536  
Db 439 LVIEFWVTFKVPETGRTEETITRAFEQVQVGTGRKEG 478

RESULT 14  
US-09-299-549-5  
; Sequence 5, Application US/09299549  
; Patent No. 6136547  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Weng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/299,549  
; FILING DATE: 26-APR-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/031,392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072002  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 494 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-299-549-5

Query Match 6.0%; Score 171; DB 4; Length 494;  
Best Local Similarity 23.7%; Pred. No. 2.3e-10;  
Matches 123; Conservative 82; Mismatches 205; Indels 110; Gaps 27;  
QY 73 AVEAIGFGKFKWKLVSITGLAWMADAMEMMILSILAPOLHCEWRLPSSVQVALLTS----- 127  
Db 13 AIIATIGSFQGYN--TG---VINAPAIKDFNLTLEERSETPPSSV-LTSLWSLS 66  
QY 128 ---VVFVGMSSSTLWGNISDQYGRKTKISVLWTLFYGYLSAFAPVYSWLVL---RGL 182  
Db 67 VAIFSVGGMIGSFVGLFVNFRGRNSMLIVNLLAIAGGLMGFKIAESVEMILJRLI 126

QY 183 VGFCIGGVPOSVTLV-AEFLPMKARAKCILLIEVFWAIGTVE--VVLAVFVMPSLG--- 236  
Db 127 IGLFCGLCTGFVPMYIGEISPTALRG-----AFGLNQLGIVIGILVAQIFGLKV 176  
QY 237 -----WRWLLILSAVPLLLFAVLCFWLPESARYDVLVS-GNOEKAIATLKRI-ATENGA 287  
Db 177 ILGTEDLWPLLLGFTILPAIIQCAALPFCPESPRFLLINRKEBEKAKEILQRLWGTEDVA 236  
QY 288 ----PMPGLKLIISROEDRGKMRDLF-TPHFRWTLTLLLWFIWES-----NAFSYYGLVL 336  
Db 237 QDIQEMKDESMRS-QEKQVTVLELFRAPNYROPILIIISIMLQLSQOISGINAVYY---- 291  
QY 337 LTTLEFOAGDVCGIGISSRKKKAVEAKCSLACEYLSSEEDYMDLLLTSEFFPGVVLVTLIIDR 396  
Db 292 -STGIFKADGV-----QEPVYATIGAG-----VVNTFTVSVF-----LVER 328  
QY 397 LGRKKT-----MALCFVIFSCSLL-----FICVGRNVLTLLLFIAFAFISGG- 440  
Db 329 AGRTLHLGLGMAFCSILMTI-SLLKDNYSWMSFICIG-----AIVFVAFVFEIGPGP 383  
QY 441 ---FOAAVYVTPVPTATRALGLGTCGSMARVAGALITPFIQAQVLMLESSVYLILAVYSGC 497  
Db 384 IPWFIVAEELFGQGPRAAMAVAGCSNWTSNFLVGLLFPSS--ATFYLGAYVFIVFTVF--- 438  
QY 498 CLLAALASCLPIETKGGGLQESSHREWGOMVG-RGMHG 536  
Db 439 LVIEFWVTFKVPETGRTEETITRAFEQVQVGTGRKEG 478

RESULT 15  
US-08-928-692-11  
; Sequence 11, Application US/08928692  
; Patent No. 5958727  
; GENERAL INFORMATION:  
; APPLICANT: Brody, Howard  
; APPLICANT: Yaver, Deborah S.  
; APPLICANT: Lamsa, Michael  
; APPLICANT: Hansen, Kim  
; TITLE OF INVENTION: Methods for Modifying the Production of  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5958727 No. 5958727disk of No. 5958727th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928,692  
; FILING DATE: 12-SEPT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 4944.200-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 488 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5958727e  
US-08-928-692-11



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:40:14 ; Search time 25.88 Seconds  
(without alignments)  
1612.970 Million cell updates/sec

Title: US-09-911-667A-2  
Perfect score: 2846  
Sequence: 1 MEEDLFLQLRLPVVKFRRTG.....MVGRGMHGAGVTRNSGSQGE 548

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_68:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1280	45.0	529	T23190	hypothetical prote
2	735.5	25.8	300	S15786	glucose transport
3	734	25.8	144	T50622	hypothetical prote
4	644.5	22.6	454	T75580	probable sugar tra
5	509.5	17.9	422	G72234	hypothetical prote
6	483	17.0	400	G69757	transporter homolo
7	466.5	16.4	455	B83213	probable MFS trans
8	436	15.3	593	JC4884	organic cation tra
9	426.5	15.0	556	S50862	organic cation tra
10	409	14.0	742	A43344	synaptic vesicle p
11	403	14.2	742	S27263	synaptic vesicle p
12	402.5	14.1	683	S34961	transmembrane tran
13	395	13.9	724	S50531	probable MFS trans
14	381.5	13.4	452	B83033	probable MFS trans
15	366	12.9	446	B83033	probable MFS trans
16	357.5	12.6	557	JW0089	organic cation tra
17	355	12.5	448	G83616	4-hydroxybenzoate
18	335.5	11.8	557	JE0346	high-affinity earn
19	334.5	11.7	576	T22509	hypothetical prote
20	333.5	11.7	423	S74046	probable sugar tra
21	333.5	11.7	448	H83335	probable MFS trans
22	322	11.3	451	S66008	transport protein
23	321.5	11.3	444	E83033	probable MFS trans
24	311	10.9	457	E70070	metabolite transpo
25	307	10.8	423	A83383	probable MFS trans
26	297	10.4	401	E69501	sugar transporter
27	296	10.4	450	C83394	probable MFS trans
28	295.5	10.4	409	T47026	hypothetical prote
29	291	10.2	469	G65058	hypothetical prote

30	287.5	10.1	442	2	A83122	probable MFS trans
31	286.5	10.1	443	2	E64725	yaaU protein - Esc
32	285.5	10.0	443	2	H85485	probable transport
33	280.5	9.9	407	2	C64167	hypothetical prote
34	279.5	9.8	418	2	A64763	probable transport
35	279.5	9.8	418	2	D85530	probable transport
36	275	9.7	557	2	T38125	myo-inositol trans
37	267	9.4	464	2	F69587	L-arabinose transp
38	266	9.3	539	2	C96758	probable protein
39	265.5	9.3	473	2	G69789	sugar transport
40	265.5	9.3	794	2	T27870	hypothetical prote
41	265	9.3	511	2	A84537	probable sugar tra
42	264.5	9.3	547	2	A48442	membrane transport
43	263	9.2	522	2	E86246	glucose transport
44	262	9.2	522	2	S12042	glucose transport
45	260.5	9.2	480	2	T23608	hypothetical prote

ALIGNMENTS

RESULT 1

T23190

hypothetical protein ZK637.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Jun-2000

C;Accession: T23190

R;Craxton, M.

submitted to the EMBL Data Library, April 1993

A;Reference number: Z19704

A;Accession: T23190

A;Status: preliminary; translated from GB//MBL/DBD

A;Molecule type: DNA

A;Residues: 1-529 <WIL>

A;Cross-references: EMBL:Z22175; PIDN:CAA80131.1; GSPDB:GN00021; CESP:ZK637.1

A;Experimental source: clone K01F9

C;Genetics:

A;Gene: CESP:ZK637.1

A;Map position: 3

A;Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3; 422/2; 472/3

C;Superfamily: Caenorhabditis elegans glucose transport protein

Query Match 45.0%; Score 1280; DB 2; Length 529;

Best Local Similarity 52.8%; Pred. No. 1.3e-97;

Matches 258; Conservative 74; Mismatches 125; Indels 32; Gaps 7;

Qy	42	HVG---	LEAVELDDGAAPKPEFANPTDD-----TFMVEDAVEAIGFGFOWKLSVLT	90
Db	34	HVGDDFAVRYNLD	DT-----ELGPTDQRSPDSEKFTTVDEAVEALGFGFQKLKLSLT	89
Qy	91	GLANMADAMEMMILS	LAPOLHCEWRPLPSWQVALLTSYVFVGMSSSTLMGNI	150
Db	90	GMANMADAMEMMILS	LIPALACENGISVQALVTCVFGSMMLSTFWGKICDRFCRR	149
Qy	151	TGLKISLWTLTYG	LISAFAPYSHWLVRLGLVGFVGIGVQSVTLYAEFLPMKAKCI	210
Db	150	KGLFTSTLVACIM	GVMSHPHYVLLFFRGLTGGVGVQSVTLYAEFLPTQRAKCV	209
Qy	211	LLIEVFVAIGTVF	VEVAVFVMPSLGWRWLLILSAVPLLFVLCFWLPESARYDVLSGN	270
Db	210	VLIESFAIGAVF	EALLAYFVNESFGWRALMFLSLPLGIFAVASFVLPESARFDMASGH	269
Qy	271	QEKATLTKRIAT	ENGAMPGLKLIISRQ----EDRKKMRDLFTPHFRWTTLLLPFIWFSN	327
Db	270	PERALETLQAAAR	NRVQLPTGRVLSVSTKAGSESGDIANLLSPDLRKTTILLMCIWAIT	329
Qy	328	AFSYGYGLVLLT	EFOAGDVC--GISSRKKAVEAKCSLACEVLSSEDMYLLTWTLSRFP	385
Db	330	AFSYGMVWLT	FTVLVFSQSHDECHGLFSNGTQME-----VCQPLTRSDTFDLSLTLAEFP	384
Qy	386	GVLVTLTIIDRL	GRKKTMLACFVIFSFCSLLIFTCVGRNVLTLLLFIAFIATISGGFQAA	445

Wed Mar 13 13:08:14 2002

Db 385 GLIITVLIIEFGRKKTMALEYAVFAIFTLFLYFCLDRFTVTVLIFVARAFISGAFOQAY 444  
 QY 446 VYTPVYPTATRALGLGTCGSMARVALITPFIQAQWMLSSVYLTAVYSGCCLLAALAS 505  
 Db 445 VYTPVYPTATRALGLGTCGSMARVALITPFIQAQWMLSSVYLTAVYSGCCLLAALAS 497  
 QY 506 CFPPIETKG 514  
 Db 498 LSLPIETKG 506  
 RESULT 2  
 S15786  
 glucose transport protein homolog - Caenorhabditis elegans (fragment)  
 C:Species: Caenorhabditis elegans  
 C:Date: 06-Jan-1995 #sequence\_revision 05-May-2000 #text\_change 05-May-2000  
 C:Accession: S15786  
 R:Craxton, M.; Ainscough, R.; Coulson, A.; Dear, S.; Du, Z.; Durbin, R.; Green, P.; Hall  
 Staden, R.; Sulston, J.; Thierry-Mieg, J.; Thomas, K.; Waterston, R.; Wilson, R.  
 submitted to the EMBL Data Library, May 1991  
 A:Reference number: S15786  
 A:Molecule type: DNA  
 A:Residues: 1-300 <CRA>  
 A:Cross-references: EMBL:Z11115  
 C:Genetics: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3  
 C:Introns: 31/2; 93/3; 121/3; 150/3; 192/2; 255/3  
 C:Superfamily: Caenorhabditis elegans glucose transport protein

Query Match 25.8%; Score 735.5; DB 2; Length 300;  
 Best Local Similarity 55.1%; Pred. No. 4.6e-53;  
 Matches 147; Conservative 38; Mismatches 67; Indels 15; Gaps 3;  
 QY 42 HVG-----LEAVLDDGAAYPKFANPTDD-----TFVVEDAVEAIGKQKWLKSVLT 90  
 Db 34 HVGDDFAVRYSNLDDRT----ELGEPTDQRSPDSEKTFVDEAVEALGFRQLKLSILT 89  
 QY 91 GLAWMADAMEMTILSILAPOLHCEWRPLSPQVALLTSVVFVGMSSSTLWGNISDQYGRK 150  
 Db 90 GMAWMADAMEMTILSILAPALACENWISSVQALVITCVFSCMMLSSTFWKGICDRFGR 149  
 QY 151 TGLKISVLTLYGILSAPVYVSWILVRLGLVFGIGVQVSVLYAEFLPMKARAKI 210  
 Db 150 KGLTFSLVACIMGVISGMSPHYVLLFRGLTGTGFGIGVQVSVLYAEFLPTAQRACV 209  
 QY 211 LLTEVFAICTVFEVVLAVFVMPISGLRWMLLISAVPLLLFVLCFVLPESARYDVLGN 270  
 Db 210 VLIESFAICAVFEALLAVFVMSFGWRALMFLSSLPGLIFAVASFWLPESARFDMASGH 269  
 QY 271 QEKAIATLKRIATENGAPMLGLIIS 297  
 Db 270 PERALETQAARNRVQLPTGRLVSS 296

RESULT 3  
 T50622  
 hypothetical protein DKFzp761H039.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
 C:Accession: T50622  
 R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.  
 submitted to the Protein Sequence Database, June 2000  
 A:Reference number: Z25145  
 A:Accession: T50622  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-144 <AAA>  
 A:Cross-references: EMBL:AL359592  
 A:Experimental source: adult amygdala; clone DKFzp761H039  
 C:Genetics:  
 A:Note: DKFzp761H039.1

Query Match 25.8%; Score 734; DB 2; Length 144;  
 Best Local Similarity 99.3%; Pred. No. 2.7e-53;  
 Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 405 LCFVIFSCSLLLFICVGRNVLTLLLFIAFAFISGGFQAAYVYTPVYPTATRALGLGTC 464  
 Db 1 LCFVIFSCSLLLFICVGRNVLTLLLFIAFAFISGGFQAAYVYTPVYPTATRALGLGTC 60  
 QY 465 SGMARVGAALITPFIQAQWMLSSVYLTAVYSGCCLLAALASCFPLPIETKGGGLQESSHRE 524  
 Db 61 SGMARVGAALITPFIQAQWMLSSVYLTAVYSGCCLLAALASCFPLPIETKGGGLQESSHRE 120  
 QY 525 WQEWVGRGMHAGVTRNSGSQE 548  
 Db 121 WQEWVGRGMHAGVTRNSGSQE 144  
 RESULT 4  
 F75580  
 probable sugar transporter - Deinococcus radiodurans (strain R1)  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
 C:Accession: F75580  
 R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.  
 M.; Shen, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896  
 A:Accession: F75580  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-454 <WHI>  
 A:Cross-references: GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAF12486.1; PID:964  
 C:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DRA0271  
 A:Map position: 2  
 C:Superfamily: yaaU protein

Query Match 22.6%; Score 644.5; DB 2; Length 454;  
 Best Local Similarity 30.6%; Pred. No. 2.3e-45;  
 Matches 142; Conservative 95; Mismatches 190; Indels 37; Gaps 5;  
 QY 70 VEDAVEAIGFGRFQKLSVLTGLAWMADAMEMTILSILAPOLHCEWRPLSPQVAV--LLTS 127  
 Db 15 VDRVDDDLGLGRFQKLLAICGLTWAADAMEVLLMGFALPGTSAAFELPKGSPAATMLLT 74  
 QY 128 VYFVGMSSSTLWGNISDQYGRKTKGLKISVLTLYGILSAPVYVSWILVRLGLVFGIG 187  
 Db 75 ATFAGMLFGAMFGYLAADRVGRSVLTTVALGVVFGLAGALAPTLTLLVAVRFLTGFAI 134  
 QY 188 GG-VPOSVTLVYAEFLPMKARAKICILLIEVFAIGTVFVFLAVFV----MPSLGRWRLLI 242  
 Db 135 GGTLPVDYVSMMAEFVPTAWRGFLVYVLESFVAVGTVVAAALAWWYSTAFAPAEGRWLLIG 194  
 QY 243 LSAVPLLLFVLCFVLPESARYDVLGSGNOKAIAITLKRIATENGAPMLGLIISQEDR 302  
 Db 195 LAALPGLVGLIARIIGTPDPSRSLARGEAQAALQVQAQANGGTLPAAPLAHPEQPPR 254  
 QY 303 GKMRDLTFPHFRMTTLLWFIWFSNAFVYGLV-----LLTTELFQAGDVCGISSRKAV 357  
 Db 255 VSPAQLFRGVLAARRTPLLVMYVWGLSLGYGIFSWLPFLRAQGLDLGAV----- 304  
 QY 358 EAKCSLACEYLSEEDYMDLLWTLTLEFFQVLTLMIDRLGRKKTMALCFVIFSCSLLL 417  
 Db 305 -----YRSTLLALAQVPGYLLAAVLEKIGRRVTLVGLFTLGAAGVAYLF 349  
 QY 418 FICVGRNVLTLLLFIAFAFISGGFQAAYVYTPVYPTATRALGLGTCGSMARVALITP 477  
 Db 350 LLAHDANTVLTLSALLSPALLGANGSLYATPPELFTPLRTTGMGLVSGVARLASVVSFS 409

QY 478 IAOVMLESSVYLTLAVYSGCLLAALASCEFLPIETKGGQLQESS 521  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 410 IGAMLLTGNLTALTFAVCFAALAALAAWIGVETRGOALAEYA 453

RESULT 5  
G72234  
hypothetical protein TM1603 - Thermotoga maritima (strain MS88)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: G72234  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickley,  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.

Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316  
A:Accession: G72234  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-422 <ARN>  
A:Cross-references: GB:AE001804; GB:AE00512; NID:g4982160; PIDN:AAD36670.1; PID:g4982172  
A:Experimental source: strain MSB  
C:Genetics:  
A:Gene: TM1603  
C:Superfamily: hypothetical protein H1104

Query Match 17.9%; Score 509.5; DB 2; Length 422;  
Best Local Similarity 29.5%; Pred. No. 2.7e-34;  
Matches 135; Conservative 81; Mismatches 188; Indels 53; Gaps 11;

QY 70 VEDAVEATGCGKQWKLSVLGTGLAWDAMEMMILSIAPOLHCEWRLPSSQVALLTSVV 129  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 3 IDEIVEKYVDRTORRFILITSIAWMFDAAGVLLSFVLPVIKEWNLTSTQGATIASAT 62

QY 130 FVGMMSSSTLWGNSDQYRGTKLGKISLVLTLYYGIIISAFAPVYSWIILVRGLGVFGIG 189  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 63 FLGMLFGALSVCVFADLLGRKVSNLFIFVITITFFLSGSSSFETLLVRLGLSGFYGG 122

QY 190 VPQSUTLY-AEFLPMKARAKICILLIEVFAIGTVFVEVVLVFMPSLGRWRLILLSAVPL 248  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 123 LMPFSNAVLAFTSIRLGRYLVLLSESSAVGSILIGLFAVNVLPN--WRWFVFIISG- 179

QY 249 LLFAVLCFWLPESARYDVLSNGOEKAIALI-KRIATENGAP----MPLGKLILISROEDRG 303  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 180 YLFVPVFLRMPTPKYAFALKGGEKALENSLGRVVEEVVELPKKKVPIALL-----KRE 234

QY 304 KMRDLTFPHFRWTILLWFIFWSNAFSYGLVLTTELFOAGDVCGISRRKKAWEAKCSL 363  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 235 HLKD-----TWVIWIANFVSFVYYLFTWAPRIFS---LGVSVKSS----- 275

QY 364 ACEYLSEEDYMDLLWTT----LSPEFGVLVTLIIDRLGRKKTWALCFVIFSCSLLLFI 419  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 276 -----WFTPYMVVAQLPGYLSAAYFIEKWGRKASLGVPYFIGTGLAALLWAN 321

QY 420 CVGRNVLTLLLFIARAFISGGFQAAYVTVTPEYVTPATRALGLTGCSMGARGVALITPFIA 479  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 322 VRGDASLLAAALVLSFFCLGVWGLVYAYTPELYPTSLRGTCNGAAGVWARIATIPIYT 381

QY 480 QVMLE-SSVYLTAVYSGCLLAALASCEFLPIETKG 514  
| : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 382 GFMEKGRKSIAETLAWISAMAFAGVIVLIFGRETKG 418

RESULT 6  
G69757  
transporter homolog yceI - Bacillus subtilis  
C:Species: Bacillus subtilis  
C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
C:Accession: G69757  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertelli,

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.D.; Emmerson, P.F.; Entlan, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997

A:Authors: Foulger, D.R.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Hollisappel, S.; Hosono, S.; Hullo, M. Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, K.; Tosato, V.; Uchiya tuchi, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

A:Reference number: A69580; MUID:198044033

A:Accession: C69757

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-400 <X>

A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12089.1; PID:g26325 A:Experimental source: strain 168

C:Genetics:

A:Gene: yceI

C:Superfamily: hypothetical protein H1104

Query Match 17.0%; Score 483; DB 2; Length 400;

Best Local Similarity 30.7%; Pred. No. 3.9e-32;

Matches 137; Conservative 75; Mismatches 160; Indels 74; Gaps 14;

QY 83 QWKLVSITGLAWADAMEMMILSLAPOLHCIEWKLPNSQVALLTSVYFVGMSSSTLWGN 142

DB 9 QKLLGVAGLGNLFDANDVGLTSLFIIAHLVHWNLSPEEMKKGVSYSIGMAAGAFILGL 68

QY 143 ISDQYGRKTGLKISVLWTLYYGI---LSAFAPYVSWILVRLGLVGFQIGG-VPOSVTLYA 198

DB 69 LADRIGRK---KVFIIITLLCFSGISAFVTSLSAFLILRFVIGMGLGGLPVAFTLS 125

QY 199 EFLPMKARAKILLIEVFWAIGTGVFVAVFVMPSLGWRWLLLSLAVPLLFAVLGFWL 258

DB 126 EAVVPEKLCRVIVLLESFWAGWGLAALISYFVTPSGWQAALLLTALTAFALYALRTSL 185

QY 259 PESARYDVLNQEKATATLKRIATENGAPMLGKLLISQREDGRKMRDLTPHFRTTL 318

DB 186 PDSPTYESLSAK-----KRSWMEN-----VKSVMARQYIR-----PTV 218

QY 319 LLWFTWFSNAFSYCYGLVLLTTTLFQAGDVCGISRRKKAWEAKCSLACEYLSEEDYMDLLW 378

DB 219 MLSIVMFCVFYSYGMFLWLPVSLMLKGFMSIQS-----FEY-----VLL 258

QY 379 TTLSEFFGLVLTLMIIIDLGRK-----KTMALCFVFSFCSLLLLFCVGRNVLLTLL 430

DB 259 MTLAQLPGYFAAMLIEKAGRKWILVYVLIGTAGSAFFGCTADSLSLTAG-----VLLS 314

QY 431 FIARAFISGGFQAAYVTPPEYVTPATRALGLTCSGMARVAGALITPFIAQVMLESSVLT 490

DB 315 F-----FNUGMGVLYAYTPQYPTAIRATGSGTTAAFRIGGIFGPLLVGLTGAARHSFS 370

QY 491 LAVYSGCCLLAALA-SCFLPI--ETK 513

DB 371 V-IFSIFCIALLAVACILINGKETK 395

RESULT 7

BB3213

probable MFS transporter PA3467 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C:Species: *Pseudomonas aeruginosa*

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: BB3213

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: AB2950; MUID:20437337  
A:Accession: B83213  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-455 <STO>  
A:Cross-references: GB:AE004767; GB:AE004091; NID:g9949604; PIDN:AAG06855.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
C:Gene: PA3467  
C:Superfamily: hypothetical protein b1775

Query Match 16.4%; Score 466.5; DB 2; Length 455;  
Best Local Similarity 27.7%; Pred. No. 1e-30; Mismatches 189; Indels 61; Gaps 13;  
Matches 132; Conservative 94; Mismatches 189; Indels 61; Gaps 13;  
  
QY 72 DAVEAIGFGKFWKLSVITGLAWMADAMEMTILSLAPOLHCEWRLPSSQVALLTSVVVF 131  
DB 10 ERLERPLSPYHRLVFVIALAFFDSDMLAMTFLGSIKAEFGLDLSAQAGLLASSSFF 69  
QY 132 GMMSSSTINGNISDOYGRKTKGKIS-VLWTLYYGILSAFAPYVSWILVRLGLVGFIG-G 189  
DB 70 GMYGAALSGMLADRFGRKPVQASIVLWGLASYLCSTAGDLS-LTFYRVLLGICMGME 128  
QY 190 VPOSVTLYAEFLPMKARAKCILLIEVFWAIGTVFVWLVAVFVWPSLGNRWLLILSAVPLL 249  
DB 129 FPTAQLSEMIPASRGKYIALMGDFWPLGFVAACLSYFLPLTGTWRSIFLVLPALPAV 188  
QY 250 LFAVLCFWLPESARYDVLGNOEKAATLKRATPE-----NGAMPGLKLIISROEDR 302  
DB 189 FVLAIRFLIPSPRWLEQAGRQEDRVLRDIARVMRSLGTELPPL-----RQPOR 242  
QY 303 GKMR-----DLTPHFRWTLTLWTFESNAFSYVGLVLLTTLFQAGDVCIGSSRK 354  
DB 243 ERSRPGFFAFAPLWSPAYRRTLTWGLFWFALLGFGYGLTSLWLSALLQSGFA----- 296  
QY 355 KAVEAKCSLACEYLSSEDDYMDLLTTLSEFPGLVTLWIIDRLGRKKTWALCF-----VI 409  
DB 297 -----VTQSVYTVTL-ISLAGIPGFLCAAWLVESMGKRPSCVLMLLGGGAMA 342  
QY 410 FSFCSLLLFICVNRVTLTLF--IARAFISGGFQAAYVTPVPTATRALGLGFCGCM 467  
DB 343 YATGQAVF---GGSLALLIGFLGAMQFFLFGWAVLYTYTPELYTSARATSGSFASAV 399  
QY 468 ARVGALITPFIQAVML-----ESSVYTLAVYSGCCLLAALASCLFPIETKGGGLQE 519  
DB 400 GRIGSLGLPLVTGLVPLTGQGGVFTLGAL---CFGVAALVWNAFGIETRGRTLEE 452  
  
RESULT 8  
JC4884  
Organic cation transporter protein 2 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 10-Sep-1996 #sequence\_revision 18-Oct-1996 #text\_change 05-Nov-1999  
R:Okuda, M.; Saito, H.; Urakami, Y.; Takano, M.; Inui, K.  
Biochem. Biophys. Res. Commun. 224, 500-507, 1996  
A:Title: cDNA cloning and functional expression of a novel rat kidney organic cation transporter  
A:Reference number: JC4884; MUID:96295517  
A:Accession: JC4884  
A:Molecule type: mRNA  
A:Residues: 1-593 <OKU>  
A:Cross-references: DDBJ:D83044; NID:g1502282; PIDN:BAAL1754.1; PID:dl012421; PID:gl5022  
A:Experimental source: kidney  
C:Comment: This protein is responsible for the transport of cationic drugs in kidney.

Query Match 15.3%; Score 436; DB 2; Length 593;  
Best Local Similarity 29.5%; Pred. No. 4.4e-28;  
Matches 127; Conservative 77; Mismatches 148; Indels 78; Gaps 16;  
  
QY 119 SNOVALTTSVVFVGMSSSTLWGNISDOYGRKTKGLKISVLTWLYYGLSFAFAPYVSWILV 178

DB 146 SMLDLDFOSVNVGFFICAMMIGYLADRFGRKFCLLTILINAIISGALMAISPNTAMLV 205  
QY 179 ---LRGLV---GFGIGGVPSQSVTLYAEFLPMKARAKCILLIEVFWAIGTVFVFWLA--VF 230  
DB 206 FRFLOGLVSKAGWLIGYI-----LITEFVGLGYRRMVGICYOIAFTVGL---LILAGVAY 257  
QY 231 VMPSLGNRWLLTSLAVPULLLFAVLGFW-LPESARYDVLGNOEKAATLKRATENGAPM 289  
DB 258 VIPN--WRWLQFAVTLPNFCF-LLYFWCIPESPRWLISONKIVKAMKIIKHIAKNGKSV 314  
QY 290 PLGKLIISROEDRGK-----MRDLF-TPHFRWTLTLWTFESNAFSYVGLVLLTTLFQ 343  
DB 315 PVSLOQLTTPDEDAGRKLKPSILDLVTRTPQIRKHTLILMYNWFSTSSVLYOGLIN---HMLG 371  
QY 344 AGDVCGISRRKKAKEACSLACEYLSSEDDYMDLLTTLSEFPGLVTLWIIDRLGRKKTW 403  
DB 372 AGD-----NYLDFFSYVSAFPAAFIILITIDRGRRPYW 407  
QY 404 ALC-FVIFSFCSLLLF-----CVGRNLTLLLFARAFISGGFQAAYVYTPPE 450  
DB 408 AVSNVAGAACLASFVFPDDLQWLKITTACLRMGITM-----AYEMVCLVNAE 456  
QY 451 VPTATRALGLGTCGSMARVAGALITPFFAQVMLESSVYTLAVYSGCCLLAALASCLFPI 510  
DB 457 LYPTIRNLGVLVSCSMCDIGGIITPFLVYRLTIWMEPPLVVFVAVGVLVAGLVLLLP- 515  
QY 511 ETKGKGLQES 520  
DB 516 ETKGKALPET 525  
  
RESULT 9  
S50862  
Organic cation transporter protein OCT1 - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 05-Nov-1999  
R:Accession: S50862; S78533; I58089  
R:Grundemann, D.; Gorboulev, V.; Gambaryan, S.; Veyhl, M.; Koepsell, H.  
Nature 372, 549-552, 1994  
A:Title: Drug excretion mediated by a new prototype of polyspecific transporter.  
A:Reference number: I58089; MUID:95082907  
A:Accession: S50862  
A:Molecule type: mRNA  
A:Residues: 1-556 <GRU>  
A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622  
A:Experimental source: kidney  
R:Gorboulev, V.G.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: S78533  
A:Accession: S78533  
A:Molecule type: mRNA  
A:Residues: 1-342, N', 344-556 <GOR>  
A:Cross-references: EMBL:X78855; NID:g633621; PIDN:CAA55411.1; PID:g633622  
C:Keywords: glycoprotein; phosphoprotein; transmembrane protein  
F:20-46/Domain: transmembrane #status predicted <TM1>  
F:154-171/Domain: transmembrane #status predicted <TM2>  
F:178-197/Domain: transmembrane #status predicted <TM3>  
F:243-260/Domain: transmembrane #status predicted <TM4>  
F:267-283/Domain: transmembrane #status predicted <TM5>  
F:350-366/Domain: transmembrane #status predicted <TM6>  
F:380-398/Domain: transmembrane #status predicted <TM7>  
F:406-425/Domain: transmembrane #status predicted <TM8>  
F:435-452/Domain: transmembrane #status predicted <TM9>  
F:469-485/Domain: transmembrane #status predicted <TM10>  
F:494-514/Domain: transmembrane #status predicted <TM11>  
F:71,97,113,432/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:286,292/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted  
F:296,343,550/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
  
Query Match 15.0%; Score 426.5; DB 2; Length 556;  
Best Local Similarity 29.3%; Pred. No. 2.5e-27;

C;Keywords: transport protein

[illegible]

Qy	337	-----LTTTELFOAGDVCG--ISSRKKAVEAKCSL-----	363
		:::	
Db	482	ARTKVFGPERVEHTFNTLENIHRGGQYFNDFKFIGLRLLKSVSFDLSLFEECYFDVTS	541
		:::	
Qy	364	-----ACEYLSEEDY-----MDLLWT-----	379
		:::	
Db	542	SNTEFRNCCTINIVFYNTDLFEYKFVNSRLVNSTFLHNKEGCPDVTGTGEGAYWYFVS	601
		:::	
Qy	380	---TLSEFPGLVLTMLIDRLGRKKKTMALCFVIPSCSLLLFCVGN---VLTLILLFIA	433
		:::	
Db	602	FLGTAVLPNGVISALLMDXIGRLMLAGSSVL--SCVSCFFILSFSGNESAMIALICLFG	659
		:::	
Qy	434	RAFISGGFOAAYYTPPEVYTATRALCLGTCSCMARVGALI-----TPFIQVMLESSVV	488
		:::	
Db	660	GVSIA-SWNALDVLTVILEYPSDKRTAFGLPNALCKLAALVGLSIFTSFVGITKAAPILF	718
		:::	
Qy	489	LTLAVSYGGCCLLAALASCFPIETKGGLQ	518
		:::	
Db	719	ASAALGSSLAKLP-----ETRGOVLQ	742
		:::	
RESULT	11		
S27263			
N;		synaptic vesicle protein SV2 - bovine	
N;		Alternate names: transporter-like protein p87	
C;		Species: Bos primigenius taurus (cattle)	
C;		Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 18-Aug-2000	
C;		Accession: S27263	
R;		Gingrich, J.A.; Andersen, P.H.; Tiberl, M.; El Mestikawy, S.; Jorgensen,	
FEB5		Lett. 312, 115-122, 1992	
A;		Title: Identification, characterization, and molecular cloning of a novel	
A;		Reference number: S27263; MUID:93050176	
A;		Accession: S27263	
A;		Molecule type: mRNA	
A;		Residues: 1-742 <GN>	
A;		Cross-references: PID:N:q259173; PIDN:AAB24028.1; PID:g259174	
C;		Superfamily: synaptic vesicle protein SV2	

and, in fact, a brain synaptic vesicle protein homologous to bacterial transporters.

A;Reference number: A43267; MUID:92390722  
A;Accession: A43267

A;Molecule type: mRNA

A/Residues: I-339, 'F', 341-742 <BAJ>  
A/Cross-references: GB:L05435; NID:q207091; PIDN:AAA42188.1; PID:q207092

A;Experimental source: brain

n; note. sequence extracted from NCBI database (NCBI:112489)  
R;Gingrich, J.A.; Andersen, P.H.; Tiberi, M.; El Mestikawy, S.; Jorgensen, P.N.; Fre

FEBS Lett. 312, 115-122, 1992

A; Reference number: S27263; MUID:93050176

A;Accession: A58948  
A;Status: nucleic acid sequence not shown

A; Molecule type: mRNA  
A; Residues: 1-120 / 5' 122-240 / 3' 251-604 / 5' 606-743 / 3' 754-1029

A, residues: 1-120, E, 122-249, W, 251-684, P, 686-742 <GIN>  
A; Note: only differences from the bovine translation are shown in Fig. 2

C;Superfamily: synaptic vesicle protein SV2





C;Superfamily: synaptic vesicle protein SV2

Query Match 13.9% Score 395; DB 2; Length 724;  
Best Local Similarity 22.4%; Pred. No. 1.3e-24;  
Matches 145; Conservative 96; Mismatches 218; Indels 188; Gaps 19;  
Qy 21 ESARSEDTSAGEHVOI-EGVHVGL-----EAVELDDGAAPKPF-----AN 62  
Db 68 EDEGSSDATEHDEDEIEYEGYQIPDMSOKKESQVAIGVLSDEYKDEELDAERRAD 127  
Qy 63 PTDDTFWVEDAVEALGCKGKFWKLSVLTGLANMADAMEMMILSLAPOLHCEWRLPSQV 122  
Db 128 EEEAQAQVELIIOCGHGRFOWALFLVGLSLMADGVEFVVGVLPSAETDMCVENSNS 187  
Qy 123 ALLTSVFPVGMSSSTLWGNISDOYGRKTKLISVLMTLYYGLSFAFVYVSWILVLRGL 182  
Db 188 GWLGSIVYLGMLGAFFWGLGADKMRQRTLIICMSINGPFAFLSSFGVGSFLFCRFF 247  
Qy 193 VGFGLGG-VPOSVTLVAFPLPKARAKCILLIEVFWALGIVFEVVLAVFVWPSLG----- 236  
Db 248 AGFGIGGAVPVVFAFVFAVTLAREKRGHLSWLCMFWMIGGIYASAMAWAIIPHYGWSFSM 307  
Qy 237 -----WRWLLILSAVPLLLFVLCFWLPESARYDVLGSGNQEKAIA-TLKRIATENGAPMLGKLIISROEDRGRMDLFT 285  
Db 308 GSAYQFHSRVFVVCALPCSSVVALTFMPESPRLYLEVKGKHDENMILKQIHDTNMRA 367  
Qy 286 -GAP-----MPLGKLI-----ISROEDRGRMDLFTPHFRW----- 315  
Db 368 RGOPEKVFVTRIKTPKLIDELIEIQDTDG-----TWYMRWFVRIRKTEMYGIWLTFRKC 421  
Qy 316 -----TLLLWFWFSNAFSYGL----- 334  
Db 422 LDYPVKRNTILLIIVTWTLFSYGLSVWFPDVIKHLQADEYASRVKRFYGEKVEDFVN 481  
Qy 335 -----VLTTLFQ-----AGDVCIGSSRRKKAVEAKCS----- 362  
Db 482 FTLENOIHTNGEYIRDRETFIMKFAVTFEDSLFNKCYFEDITSLSYFK-----NCTFTET 537  
Qy 363 -----LACEYLS-----EEDYMDLLWT-----TLSEEPGVLV 389  
Db 538 LFYNTDLSEEFIDCQFNSTFLHNKKGCCINFDEY-SAYVIYFVNPLGTALVPLGNIV 596  
Qy 390 TLWIIDRLGRKKTALCFVIFSCSLLIFCVGRNVLTLFLFIARAFISGFGQAAVYVTP 449  
Db 597 SALLMDRIGRLTMLGSGSVLSCISCFELWFGTSEAMMIGMLCYNGLTISAWNSLDVTV 656  
Qy 450 EYVPTATRALGLG---TCSGMARVGLITPFIQVWLESVYLTLA 492  
Db 657 ELLPTDRRATGFGFLNALCKAATVGNLI--FGSLVGLITKSIPIMLA 701  
RESULT 14  
A85852  
Probable transporter Z3394 [imported] - Escherichia coli (strain O157:H7)  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001  
C;Accession: A85852  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouasis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: A85852  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-452 <STO>  
A;Cross-references: GB:AE005174; NID:g12516456; PIDN:AAG57277.1; GSPDB:GN00145; UWGP:Z33  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z3394

Query Match 13.4% Score 381.5; DB 2; Length 452;  
Best Local Similarity 25.3%; Pred. No. 1e-23;  
Matches 113; Conservative 88; Mismatches 194; Indels 51; Gaps 10;  
Qy 74 VEATGEGKFWKLSVLTGLANMADAMEMMILSLAPOLHCEWRLPSQVVALTTSVVFVGM 133  
Db 11 IDAAPVGMQWRVVICCFVLVWLDGFDTAAGFIAPDIRTHWQLTAGDLSPLFGAGLLXL 70  
Qy 134 MSSSTLWGNISDOYGRKTKLISVLMTLYYGLSFAFVYVSWILVLRGLVGRGIGG-VPO 192  
Db 71 TAGALLCPLSDRFGKRVIELCVFLGALSASFDPDLQTLVFLFLTGLGLGAMPN 130  
Qy 193 SVTLVYAEPLPKARAKCILLIEVFWALGIVFEVVLAVFVWPSLGWRLMILSLAV-PLLLF 251  
Db 131 TITTSYELPARRRGALVTLMPGFTLGSAGFIVSAOLVPPVIGHGILVGLGVLPMLFL 190  
Qy 252 AVLFCFWLPESARYDVLGSGNQEKAIA-TLKRIATENGAPMLGKLIISROEDRGRMDLFT 310  
Db 191 VALLVLPESPRMQRVROLPOAVIAKTVSAITRERYVDTHF-YLIESASVTKGSIRQLFM 249  
Qy 311 PHFRWTTLLMLPFIWPSNFAFSYGLVLLTTELPQADGVCIGISSRKKAVEAKCSLACEYLSE 370  
Db 250 GROLPIITLMLWVVFMSLLIY---LLSSWMTLLNHRGI----- 286  
Qy 371 EYMDLLWTTLS-BFPQVGLVTL---WIIDRLGRKKTALCFVIFSCSLLIFCV-GRNV 425  
Db 287 -DLQASWVTAAFQGGTGLGALGVLMDKNPFVRLTSLYAIGAICIVMIGLSUGLWL 345  
Qy 426 LTLTLFIARAFISGFGQAAVYVTPVPTATRALGTCGSMARVGLITPFIQVWLESVYLT 484  
Db 346 MALAIFGTGIGTSQVGLNALTATLYPTOSRATGVSWSNAIGRCGAIVGSLSGGVMMAM 405  
Qy 485 -----SSVYLTAV 493  
Db 406 NFSDTLFFIITAVPAASAVMLTLII 431  
RESULT 15  
B83033  
Probable MFS transporter PA4900 [imported] - Pseudomonas aeruginosa (strain PA01)  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C;Accession: B83033  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A;Reference number: A82950; MUID:20437337  
A;Accession: B83033  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-446 <STO>  
A;Cross-references: GB:AE004903; GB:AE004091; NID:g9951173; PIDN:AAG08285.1; GSPDB:GN  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA4900  
C;Superfamily: yaaU protein  
Query Match 12.9% Score 366; DB 2; Length 446;  
Best Local Similarity 24.8%; Pred. No. 1.9e-22;  
Matches 104; Conservative 79; Mismatches 197; Indels 40; Gaps 7;  
Qy 74 VEATGEGKFWKLSVLTGLAW-----MADAMEMMILSLAPOLHCEWRLPSQVVALTTSV 128  
Db 10 IDAARFNTRHLLI-----LGNGCFIMLFDGVDWIVGVSVVPRMQEWQLSPVQAGTLGSC 64  
Qy 129 VFGVMSSSTLWGNISDOYGRKTKLISVLMTLYYGLSFAFVYVSWILVLRGLVGLV 184  
Db 65 ALFGMLFGGTLTALPLADFRGRRR-----LVIAITLLASLAFLTGHDAPLELGAGRFTFTG 120  
Qy 185 FGIGG-VPOSVTLVYAEPLPKARAKCILLIEVFWALGIVFEVVLAVFVWPSLGWRLMIL 243

Db 121 LAUGALVPSAINLISEFAPAGRRSTLTWNSAFYSVGAVLSALLATAMIPANGWQSVFYV 180  
QY 244 SAVPLLLFAVLCFWLPESARYDVLSCNQEKATATLKRIATENGAPMPLGKLIISROEDRG 303  
Db 181 AVLPVLAVPLMLRWLPESAAFLLELKGRRAELDALLRKVDPDY-RPGAERANAVAAEAPSG 239  
QY 304 KMRDLFTPHFRWTTLLWFIWFSNAPFSYYGLVLLTTTELFOAGDVCGIISRKKAVEAKCSL 363  
Db 240 RVAQLFEGRQAVGTLLLLWAFAMCMLMSYGLNTWLPKLMAGGG----- 282  
QY 364 ACEYLSSEEDYMDLWTTLTSEFFGVLTLMIIIDRLGRKKTMAICFVIFSCSULLLFCVGR 423  
Db 283 ---YALGSSLAFVLTNVGATLGALFGGWLADRLGAGRTLVLFFALAAASLAAALGLGPGP 339  
QY 424 NVLTLLFIARAFISGGFOAAVYVTPPTATRALGLGTCSGMARVGALITPFIQVVML 483  
Db 340 WLLNGLLVVAGATTIGTLAVIHAYAAQFYPAWVRSTGVGWAAGVGRGAIAGPMLGGSLL 399

Search completed: March 13, 2002, 12:41:38  
Job time: 84 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:40:14 ; Search time 17.05 Seconds  
(without alignments)  
1178.435 Million cell updates/sec

Title: US-09-911-667A-2  
Perfect score: 2846  
Sequence: 1 MEEDLFQLRLPVVKFRRTG.....MVGRMGHAGVTRNSGSQ 548

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Sw:ssProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1280	45.0	529	Y0U1_CAEEL	P30638 caenorhabdi
2	483	17.0	400	YCE1_BACSU	O34691 bacillus su
3	408	14.3	742	SVY2_RAT	Q02563 rattus norv
4	357.5	12.6	557	OCN2_HUMAN	O76082 homo sapien
5	353	12.4	448	PCAK_PSEPU	Q51955 pseudomonas
6	346.5	12.2	457	PCAK_ACICA	Q43975 acinetobact
7	339.5	11.9	557	OCN2_RAT	O70594 rattus norv
8	332	11.7	557	OCN2_MOUSE	Q92068 mus musculu
9	322	11.3	451	Y0U1_CAEEL	P37514 bacillus su
10	295.5	10.4	413	MUCK_ACICA	P94131 acinetobact
11	295.5	10.4	466	BENK_ACICA	O30513 acinetobact
12	286.5	10.1	443	Y0U1_CAEEL	P31679 escherichia
13	284	10.0	751	Y0U1_CAEEL	P46501 caenorhabdi
14	282.5	9.9	445	Y0U1_CAEEL	P46501 caenorhabdi
15	280.5	9.9	407	Y0U1_CAEEL	P46501 caenorhabdi
16	279.5	9.8	403	Y0U1_CAEEL	P46501 caenorhabdi
17	275	9.7	557	Y0U1_CAEEL	P46501 caenorhabdi
18	264.5	9.3	547	Y0U1_CAEEL	P46501 caenorhabdi
19	262	9.2	522	Y0U1_CAEEL	P46501 caenorhabdi
20	260.5	9.2	428	Y0U1_CAEEL	P46501 caenorhabdi
21	255	9.0	541	Y0U1_CAEEL	P46501 caenorhabdi
22	254	8.9	575	Y0U1_CAEEL	P46501 caenorhabdi
23	249.5	8.8	473	Y0U1_CAEEL	P46501 caenorhabdi
24	249.5	8.8	546	Y0U1_CAEEL	P46501 caenorhabdi
25	243.5	8.6	461	Y0U1_CAEEL	P46501 caenorhabdi
26	241.5	8.5	523	Y0U1_CAEEL	P46501 caenorhabdi
27	238.5	8.4	459	Y0U1_CAEEL	P46501 caenorhabdi
28	234.5	8.2	559	Y0U1_CAEEL	P46501 caenorhabdi
29	228	8.0	566	Y0U1_CAEEL	P46501 caenorhabdi
30	227.5	8.0	587	Y0U1_CAEEL	P46501 caenorhabdi
31	223	7.8	482	Y0U1_CAEEL	P46501 caenorhabdi
32	221.5	7.8	464	Y0U1_CAEEL	P46501 caenorhabdi
33	221	7.8	570	Y0U1_CAEEL	P46501 caenorhabdi

34	217	7.6	1222	1	YMP3_CAEEL	Q10947 caenorhabdi
35	216	7.6	468	1	GLCP_SYNY3	P15729 synechocyst
36	216	7.6	496	1	NANT_ECOLI	P41036 escherichia
37	214.5	7.5	531	1	GHT2_SCHPO	O74969 schizosacch
38	214	7.5	567	1	HXT3_YEAST	P32466 saccharomyc
39	214	7.5	584	1	ITR1_YEAST	P30605 saccharomyc
40	213.5	7.5	592	1	HXT5_YEAST	P38695 saccharomyc
41	211	7.4	491	1	XYLE_ECOLI	P09098 escherichia
42	210	7.4	522	1	STA_RICCO	Q10710 ricinus com
43	210	7.4	535	1	GHT6_SCHPO	O74849 schizosacch
44	210	7.4	574	1	GAL2_YEAST	P13181 saccharomyc
45	208.5	7.3	452	1	YDJE_ECOLI	P38055 escherichia

ALIGNMENTS

RESULT 1						
Y0U1_CAEEL						
ID	Y0U1_CAEEL	STANDARD;	PRT;	529	AA.	
AC	P30638; Q21101;					
DT	01-APR-1993 (Rel. 25, Created)					
DT	30-MAY-2000 (Rel. 39, Last sequence update)					
DT	30-MAY-2000 (Rel. 39, Last annotation update)					
DE	HYPOTHETICAL 58.3 KDA PROTEIN ZK637.1 IN CHROMOSOME III.					
GN	ZK637.1					
OS	Caenorhabditis elegans.					
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;					
OC	Rhabditidae; Peloderinae; Caenorhabditis.					
OX	NCBI_TaxID=6239;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=BRISTOL N2;					
RX	MEDLINE=92168156; PubMed=1538779;					
RA	Sulston J., Du Z., Thomas K., Wilson R., Hillier L., Staden R.,					
RA	Halloran N., Green P., Thierry-Mieg J., Qiu L., Dear S., Coulson A.,					
RA	Craxton M., Durbin R.K., Berks M., Metzstein M., Hawkins T.,					
RA	Ainscough R., Waterston R.					
RT	"The C. elegans genome sequencing project: a beginning."					
RL	Nature 356:37-41(1992).					
RN	[2]					
RP	REVISIONS.					
RC	STRAIN=BRISTOL N2;					
RA	Durbin R.;					
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.					
CC	-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).					
CC	-I- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch).					
CC	-----					
DR	EMBL; Z11115; CAA77460.1;					
DR	EMBL; Z22175; CAA77460.1; JOINED.					
DR	EMBL; Z22175; CAA80131.1;					
DR	EMBL; Z11115; CAA80131.1; JOINED.					
DR	PIR; S15786; S15786.					
DR	WormPep; ZK637.1; CE06638.					
DR	InterPro; IPR003662; sub_transportr.					
DR	Pfam; PF00083; sugar_tr; 1.					
DR	PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.					
DR	PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.					
KW	Hypothetical protein; Transmembrane; Transport.					
FT	TRANSMEM 86 106					
FT	TRANSMEM 122 142					
FT	TRANSMEM 158 178					
FT	TRANSMEM 238 258					
FT	TRANSMEM 320 340					
FT	TRANSMEM 373 393					

```
FT TRANSMEM 411 431 POTENTIAL.
FT TRANSMEM 482 502 POTENTIAL.
SQ SEQUENCE 529 AA: 58317 MW: 8D2FF4CBA15ECD2D CRC64;

Query Match 45.0%; Score 1280; DB 1; Length 529;
Best Local Similarity 52.8%; Pred. No. 5.1e-84;
Matches 258; Conservative 74; Mismatches 125; Indels 32; Gaps 7;

QY 42 HVG-----LEAVELDDGAAPREFANPTDD-----TFWEDAVAEIGFGRFOWKLSVLT 90
DB 34 HVGDDFAVRYSNLDR-----ELGPTDQSPDSEKTTTVDVEAEALGGRFQLKLSILT 89
QY 91 GLAWMADAMEMMILSILAPOLHCEWRLPSWQVALLTSVVFVGMSSSSTLWGNISDOYGRK 150
DB 90 GMAWADAMEMMILSILPALACENGISVQALVTTCTVFGSMMLSSSTFWGKICDRFGR 149
QY 151 TGLKISVLWLTLYGILSAFAPVSYWILVRLGVLGFGIGVQSVTLVYAEFLPMKARAKCI 210
DB 150 KGLTSTLVACIMGVISCMSPHYVLLFFRGLTGFGIGVQSVTLVYAEFLPTAQAKCV 209
QY 211 LLIEVFWAIGTVFVVLAVFVWPSLWGRWMLLILSAVPLLLFAVLCFWLPESARYDVLSGN 270
DB 210 VLIESFWAIGTVFVVLAVFVWPSLWGRWMLLILSAVPLLLFAVLCFWLPESARYDVLSGN 269
QY 271 QEKATATLKRIATENGAPMPLGKLIISQ---EDRGKMRDLFTPHFRWTTLLLWIFWFSN 327
DB 270 PERALETLQAAARMRVOLPTGRVLSSTKAGSESGDITANLLSPDLRKTITLLWCIWAIT 329
QY 328 AFSYGLVLLTTELFOAGDVC---GISSRKKAWEAKCSLACEYLSSEEDYMDLLWTLSEFP 385
DB 330 AFSYGMVLTFTVLFQSHDECHGGLFSNGTOME-----VCQPLTRSDYFDLLSTLAEFP 384
QY 386 GVLVTLWIIDRLGRKKTWALCFVIESFCSLLILFCVGRNVLTLLFIARAFISGFGQAA 445
DB 385 GLIITVLLIEWGRKKTWALCFVIESFCSLLILFCVGRNVLTLLFIARAFISGFGQAA 444
QY 446 VYTPVYPTATRALGLTCSMARVAGALITPPIAQWMLSESSVYLLFVAVSGCLLAALAS 505
DB 445 VYTPVYPTATRALGLTCSMARVAGALITPPIAQWMLSESSVYLLFVAVSGCLLAALAS 497
QY 506 CFLPIETKG 514
DB 498 LSLPIETKG 506

RESULT 2
YCEL_BACSU
ID YCEL_BACSU STANDARD; PRT; 400 AA.
AC O34691;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YCEL.
OS YCEL.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunano M., Tanakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation."
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB000617; BAA22256.1; -
DR EMBL; Z99105; CABL2089.1; -
DR SUBTILIST; BG12773; ycel.
DR InterPro; IPR003662; sub_transport.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 218 238 POTENTIAL.
FT TRANSMEM 254 274 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 305 325 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
SQ SEQUENCE 400 AA: 43708 MW; E0AE0CEE5DD27395 CRC64;

Query Match 17.0%; Score 483; DB 1; Length 400;
Best Local Similarity 30.7%; Pred. No. 1.9e-27;
Matches 137; Conservative 75; Mismatches 160; Indels 74; Gaps 14;

QY 83 QNKLSTVLTLAWMADAMEMMILSILAPOLHCEWRLPSWQVALLTSVVFVGMSSSSTLWGN 142
DB 9 QNKLSTVLTLAWMADAMEMMILSILAPOLHCEWRLPSWQVALLTSVVFVGMSSSSTLWGN 142
QY 143 ISDOYGRKTGLKISVLWLTLYGCI---LSAFAPVSYWILVRLGVLGFGIGVQSVTLVY 198
DB 69 LADRIGRK---KVFTITLFCFSGISAFVTSLSAFLILREVIGMGLGGLPVAFTLS 125
QY 199 EFLPMKARAKCILLIEVFWAIGTVFVVLAVFVWPSLWGRWMLLILSAVPLLLFAVLCFWL 258
DB 126 EAVVPEKRGVIVLLEFSEWAVGNLAAALISYFVIFSGWQAALLTALTAFYALYRTSL 185
QY 259 PESARYDVLSGNQEKAIATLKRIATENGAPMPLGKLIISQ---EDRGKMRDLFTPHFRWTT 318
DB 186 PDSKYESLSAK-----KRSWMEN-----VKSVMARQYIR-----PTV 218
QY 319 LLWFIWFSNAPSYGVLVLTTELFOAGDVCIGISSRKKAWEAKCSLACEYLSSEEDYMDLLW 378
DB 219 MLSIVMFCVVFYSYGMFLWPLSVMLLKGFMSITQS-----FEY-----VLL 258
QY 379 TTLSEFPGLVLTWIIDRLGRK-----KTMALCFVIFSPFCSLLLFICVGRNVLTLL 430
DB 259 MTLAQLPGYFSAWLLIEKAGRWILVYVLTGAGSAYFFGTADSLSLLTAG-----VLLS 314
QY 431 FIARAFISGFGQAAAYVYTPVYPTATRALGLTCSMARVAGALITPPIAQWMLSESSVYIT 490
DB 315 F----FNLGAWGLVYAYTPEQYPTAIRAGSGTTAAFGRIIGGIFGLLVGTLLAARHISE 370
QY 491 LAVYSGCLLAALA-SCFLPI--ETK 513
DB 371 V-IFSIFCIAILLAVACILIMGKETK 395

RESULT 3
SYV2_RAT
ID SYV2_RAT STANDARD; PRT; 742 AA.
AC O02563;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SYNAPTIC VESICLE PROTEIN 2 (SV2).
```

GN SV2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC NCBI\_TaxID=10116;  
 CC [1]  
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.  
 RP TISSUE=Brain;  
 RC MEDLINE=92390722; PubMed=1519064;  
 RA Bajjalieh S.M., Peterson K., Shingal R., Scheller R.H.;  
 RT "SV2, a brain synaptic vesicle protein homologous to bacterial  
 RT transporters.";  
 RL Science 257:1271-1273(1992).  
 CC -!- FUNCTION: MAY ACT AS A VESICLE-LOCALIZED NEUROTRANSMITTER  
 CC TRANSPORTER OR MAY FUNCTION AS AN ION TRANSPORTER OR CHANNEL.  
 CC -!- SUBCELLULAR LOCATION: SYNAPTIC VESICLE.  
 CC -!- TISSUE SPECIFICITY: NEURAL AND ENDOCRINE CELLS OF BRAIN AND  
 CC SPINAL CORD.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L05435; AAA42188.1; ..  
 DR Pfam; PF00083; sugar\_tr; 1.  
 KW Synapse; Nerve; Glycoprotein; Neurotransmitter transport;  
 KW Transmembrane.  
 FT DOMAIN 1 163 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 164 182 POTENTIAL.  
 FT DOMAIN 183 204 LUMENAL (POTENTIAL).  
 FT TRANSMEM 205 225 POTENTIAL.  
 FT DOMAIN 226 242 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 243 260 POTENTIAL.  
 FT DOMAIN 261 262 LUMENAL (POTENTIAL).  
 FT TRANSMEM 263 281 POTENTIAL.  
 FT DOMAIN 282 294 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 295 322 POTENTIAL.  
 FT DOMAIN 323 334 LUMENAL (POTENTIAL).  
 FT TRANSMEM 335 355 POTENTIAL.  
 FT DOMAIN 356 445 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 446 469 POTENTIAL.  
 FT DOMAIN 470 592 LUMENAL (POTENTIAL).  
 FT TRANSMEM 593 611 POTENTIAL.  
 FT DOMAIN 612 626 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 627 647 POTENTIAL.  
 FT DOMAIN 648 649 LUMENAL (POTENTIAL).  
 FT TRANSMEM 650 669 POTENTIAL.  
 FT DOMAIN 670 694 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 695 711 POTENTIAL.  
 FT DOMAIN 712 712 LUMENAL (POTENTIAL).  
 FT TRANSMEM 713 731 POTENTIAL.  
 FT DOMAIN 732 742 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 742 AA; 82705 MW; 565DE7EF2929D5DB CRC64;

Query Match 14.3%; Score 408; DB 1; Length 742;  
 Best Local Similarity 21.9%; Pred. No. 7.9e-22;  
 Matches 151; Conservative 101; Mismatches 232; Indels 206; Gaps 20;  
 QY 20 GEGARSED-----DTASG--EHEVQIEGVHVGLEAVE-----LD 51  
 Db 68 GEGAQDEEGCASSDATEGHEDEDEIYEYQGIIPRASEGKGGERMADGAPLAVRGSL 127  
 QY 52 DGAAVP-----KEFANPTDDTFMVEDAVEAIGFGKFWKLSVITGLAWADAM 99  
 Db 128 DGEPPGGGRGEAQRKRDRELAQ-----QYETILRCGHGFWTLYFVLGLALMADGV 181

QY 100 EMMILSILAPOLHCEWRLPSPHOVALLTSVVFVGMSSSTLWGNISDOYGRKTKGLKISVLW 159  
 Db 182 EVFVGVFLPSAEKDMCLSDSNKGMGLIIVYGLMMVGAFLWGLADLRGRROCLLSLSV 241  
 QY 160 TLYYGILSAPAPVYVSWILVRLGLVGFGG--VPOSVTLYAEFLPMKARAKCILLIEFWA 218  
 Db 242 NSVFAFFSSVQGYGTFLFCRLLSGVGIGGSIPIVFSVSEFLAQEKGEHLSWLCMFW 301  
 QY 219 IGTVEFVVLAVFVMSLIG-----WRMLLISAVPLLLFAVLCWLPESARYD 265  
 Db 302 IGGVYAAAMAWALIIPHYGWSFGMSAYQFHSWRVFLVFAFPVSFAIGALTQPSRPF 361  
 QY 266 VLSNQEKAIATLARIATEN---GAP---MPLGKLIISROEDR----- 302  
 Db 362 LENGKHDEAMVLRKQVHDTNMRAGHPVFSVTHIKTIHQEDELIEIQSDTGTWYQRMG 421  
 QY 303 -----GKMRDLFTPHFRWTLTLLWIFSNFASYGLVL----- 336  
 Db 422 VRALSLGQVWGNFLSCFSPRYRITLMMGMVMTMSYVGLTWFPDPMIRHLQAVDYA 481  
 QY 337 -----LTTELFQAGDVCV---ISSRKAVERAKCSL----- 363  
 Db 482 ARTKVFPCRVERHVTNFNTLENQIHRGGQYFNDKFIGURLASVSFEDSLFECYFEDVTS 541  
 QY 364 -----ACEYLSEEDY-----MDLLWT----- 379  
 Db 542 SNTFFRNCFTINTVFYNTDLEFYKFNVSRLVNSTFLHNKEGCPDVTGTGEGAYMVYFVS 601  
 QY 380 ---TLSEFFPGVLVTLWIIDRLGRKTKMALCFVIFSCFLLIFCVGRN---VLTLLLFIA 433  
 Db 602 FLGTLAVLPGNIVSALLMDKIGRLMLAGSSVL--SCVSCFFLSPGNSAMIALLCFLG 659  
 QY 434 RAPTSGGFOAAVYVTPVYPTATRALGIGTCSGMARVACALI-----TPFIAQVMLESSVY 488  
 Db 660 GVSIA-SWNAIDVLITVELYPSDKRTTARFGLNACKLAALVIGISIFTSFVGITKAAPILF 718  
 QY 489 LTLAVYSGCCLLAALASCFPLPIETKGGGLQ 518  
 Db 719 ASAALGSSLALKLP-----ETRGQVLO 742  
 RESULT 4  
 OCN2\_HUMAN  
 ID OCN2\_HUMAN STANDARD; PRT; 557 AA.  
 AC 076082;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 2,  
 DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER).  
 GN SLC22A5 OR OCTN2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=98289574; PubMed=9618255;  
 RA Wu X., Prasad P.D., Leibach F.H., Ganapathy V.;  
 RT "cDNA sequence, transport function, and genomic organization of human  
 RT OCTN2, a new member of the organic cation transporter family.";  
 RL Biochem. Biophys. Res. Commun. 246:589-595(1998).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=98352077; PubMed=9685390;  
 RA Tamai I., Ohashi R., Nezu J.-I., Yabuuchi H., Oku A., Shimane M.,  
 RA Sai Y., Tsuji A.;  
 RT "Molecular and functional identification of sodium ion-dependent, high  
 RT affinity human carnitine transporter OCTN2.";  
 RL J. Biol. Chem. 273:20378-20382(1998).  
 RN [3]

RP SEQUENCE FROM N.A.  
RX MEDLINE=99113835; PubMed=9916797;  
RA Nezu J., Tamai I., Oku A., Ohashi R., Yabuuchi H., Hashimoto N.,  
RA Nikaide H., Sai Y., Koizumi A., Shoji Y., Takada G., Matsuishi T.,  
RA Yashino M., Kato H., Ohura T., Tsujimoto G., Hayakawa J., Shimane M.,  
RA Tsuji A.;  
RT "Primary systemic carnitine deficiency is caused by mutations in a  
RT gene encoding sodium ion-dependent carnitine transporter.";  
RL Nat. Genet. 21:91-94(1999).  
RN [4].  
RP CHARACTERIZATION.  
RX MEDLINE=99384224; PubMed=10454528;  
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,  
RA Chen J., Conway S.J., Ganapathy V.;  
RT "Functional characteristics and tissue distribution pattern of organic  
RT cation transporter 2 (OCTN2), an organic cation/carnitine  
RT transporter.";  
RL J. Pharmacol. Exp. Ther. 290:1482-1492(1999).  
RN [5].  
RP VARIANT CDSP GLN-169.  
RX MEDLINE=99355597; PubMed=10425211;  
RA Burwinkel B., Kreuder J., Schweitzer S., Vorgerd M., Gempel K.,  
RA Gerbitz K.-D., Killmann M.W.;  
RT "Carnitine transporter OCTN2 mutations in systemic primary carnitine  
RT deficiency: a novel Arg169Gln mutation and a recurrent Arg282Ser  
RT mutation associated with an unconventional splicing abnormality.";  
RL Biochem. Biophys. Res. Commun. 261:484-487(1999).  
RN [6].  
RP VARIANT CDSP CYS-211.  
RX MEDLINE=99408248; PubMed=10480371;  
RA Vaz F.M., Scholte H.R., Ruiter J., Hussaarts-Odiijk L.M.,  
RA Rodrigues Pereira R., Schweitzer S., de Klerk J.B.C., Waterham H.R.,  
RA Wanders R.J.A.;  
RT "Identification of two novel mutations in OCTN2 of three patients with  
RT systemic carnitine deficiency.";  
RL Hum. Genet. 105:157-161(1999).  
RN [7].  
RP VARIANT CDSP LEU-478.  
RX MEDLINE=99172075; PubMed=10072434;  
RA Tang N.L., Ganapathy V., Wu X., Hui J., Seth P., Yuen P.M.,  
RA Wanders R.J., Fok T.F., Hjelm N.M.;  
RT "Mutations of OCTN2, an organic cation/carnitine transporter, lead to  
RT deficient cellular carnitine uptake in primary carnitine deficiency.";  
RL Hum. Mol. Genet. 8:655-660(1999).  
RN [8].  
RP CHARACTERIZATION OF VARIANT CDSP LEU-478, AND MUTAGENESIS.  
RX MEDLINE=20026865; PubMed=10559218;  
RA Seth P., Wu X., Huang W., Leibach F.H., Ganapathy V.;  
RT "Mutations in novel organic cation transporter (OCTN2), an organic  
RT cation/carnitine transporter, with differential effects on the  
RT organic cation transport function and the carnitine transport  
RT function.";  
RL J. Biol. Chem. 274:33388-33392(1999).  
RN [9].  
RP VARIANTS CDSP ARG-283 AND PHE-446.  
RX MEDLINE=20081068; PubMed=10612840;  
RA Mayatepek E., Nezu J., Tamai I., Oku A., Katsura M., Shimane M.,  
RA Tsuji A.;  
RT "Two novel missense mutations of the OCTN2 gene (W283R and V446F) in a  
RT patient with primary systemic carnitine deficiency.";  
RL Hum. Mutat. 15:118-118(2000).  
RN [10].  
RP VARIANT CDSP LYS-452.  
RX MEDLINE=20145665; PubMed=10679939;  
RA Wang Y., Kelly M.A., Cowan T.M., Longo N.;  
RT "A missense mutation in the OCTN2 gene associated with residual  
RT carnitine transport activity.";  
RL Hum. Mutat. 15:238-245(2000).  
RN [11].  
RP FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE  
RP TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE  
RP INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF  
RP CARNITINE.  
RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC TISSUE SPECIFICITY: STRONGLY EXPRESSED IN KIDNEY, SKELETAL MUSCLE,  
CC HEART AND PLACENTA.  
CC DISEASE: DEFECTS IN SLC22A5 ARE THE CAUSE OF SYSTEMIC PRIMARY  
CC CARNITINE DEFICIENCY (CDSP). CDSP IS AN AUTOSOMAL RECESSIVE  
CC DISORDER OF FATTY ACID OXIDATION CAUSED BY DEFECTIVE CARNITINE  
CC TRANSPORT. PRESENT EARLY IN LIFE WITH HYPOKETOTIC HYPOGLYCEMIA AND  
CC ACUTE METABOLIC DECOMPENSATION, OR LATER IN LIFE WITH SKELETAL  
CC MYOPATHY OR CARDIOMYOPATHY.  
CC SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC  
CC CATION SUBFAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AF057164; AAC24828.1; -  
CC EMBL; AB015050; BAA29023.1; -  
CC EMBL; AB016625; BAA36712.1; -  
CC MIM; 603377; -  
CC MIM; 212140; -  
CC InterPro; IPR003662; sub\_trnsport.  
CC Pfam; PF00083; sugar\_tr; 1.  
CC PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
CC Transport; Transmembrane; Glycoprotein; Disease mutation.  
CC TRANSMEM 21 41 POTENTIAL.  
CC TRANSMEM 143 163 POTENTIAL.  
CC TRANSMEM 173 193 POTENTIAL.  
CC TRANSMEM 196 216 POTENTIAL.  
CC TRANSMEM 233 253 POTENTIAL.  
CC TRANSMEM 258 278 POTENTIAL.  
CC TRANSMEM 343 363 POTENTIAL.  
CC TRANSMEM 372 392 POTENTIAL.  
CC TRANSMEM 414 434 POTENTIAL.  
CC TRANSMEM 437 457 POTENTIAL.  
CC TRANSMEM 489 509 POTENTIAL.  
CC CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 64 64 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC CARBOHYD 91 91 N-LINKED (GLCNAC. .) (POTENTIAL).  
CC VARIANT 169 169 R -> Q (IN CDSP).  
CC VARIANT 211 211 Y -> C (IN CDSP).  
CC VARIANT 283 283 W -> R (IN CDSP).  
CC VARIANT 446 446 V -> F (IN CDSP).  
CC VARIANT 452 452 E -> K (IN CDSP).  
CC VARIANT 478 478 P -> L (IN CDSP). LOSS OF CARNITINE  
CC TRANSPORT BUT STIMULATED ORGANIC CATION  
CC TRANSPORT).  
CC MUTAGEN 352 352 M -> R: LOSS OF BOTH CARNITINE AND ORGANIC  
CC CATION TRANSPORT FUNCTIONALITIES.  
CC SEQUENCE 557 AA; 62751 MW; 928B1F6EFF63C48D CRC64;

Query Match 12.6%; Score 357.5; DB 1; Length 557;  
Best Local Similarity 24.8%; Pred. No. 2.3e-18;  
Matches 124; Conservative 96; Mismatches 204; Indels 77; Gaps 16;  
Qy 48 VELDDGAAPV---REFANPTDDTFMVEDAVEAIG-----FGKFWKLSVLTGLAMMA 96  
Db 69 LRLRDGVRPHSCRRYRLATIANF-----SALGVRGVDVLGOLEOE-SCLDGWESQ 121  
Qy 97 DAMEMMLISLAPLQHCWRL--PSQVALLTVVFGVMSSSTLWGNISDQYGRKTGL 153  
Db 122 DVLSTIVT-----EWNLCEDDWKAPLTISLFFVGVLLGSLFISQLSDRGRKNVL 173

FT	TRANSEM	95	115	POTENTIAL.
FT	TRANSEM	120	140	POTENTIAL.
FT	TRANSEM	153	173	POTENTIAL.
FT	TRANSEM	185	205	POTENTIAL.
FT	TRANSEM	262	282	POTENTIAL.
FT	TRANSEM	302	322	POTENTIAL.
FT	TRANSEM	330	350	POTENTIAL.
FT	TRANSEM	352	372	POTENTIAL.
FT	TRANSEM	399	419	POTENTIAL.
FT	TRANSEM	422	442	POTENTIAL.
SQ	SEQUENCE	448 AA;	47176 MW;	6D5IC143123E99BC CRC64;

Query Match 12.4%; Score 353; DB 1; Length 448;  
Best Local Similarity 26.6%; Pred. No. 3.9e-18;  
Matches 122; Conservative 77; Mismatches 189; Indels 70; Gaps

QY	62	NPTDDTFMVEDAVEATGFGFKFOWKLSVLTGLANMAOAMEMHILSILAQLHCERLPSMQ	121
DB	6	NSVGKSLSDVQSFINQGPLSRYSQRVLLCFLVFLDGLDTAANGFIAPALSOEWGIDRAS	65
QY	122	VALLTSVFVGMSSSTLMGNISDQYGRKTGLKSIVLUPLYGYLSAFAPVSYWILVRG	181
DB	66	LGPVMSAALIGVMFGLSGPLADRFGRKGVLVGAVLVFGGFSLASAYATNVDDLVLFR	125
QY	182	LVGFGIG-GVPQSVTYAEFLPMKARAKCTILLVEF--WAIGTVFEVWLAVFMPSLGWR	238
DB	126	LTGLGLGAGMPNATTLLSEYTP--ERUKSLVTISMFCGFNLGWAGGGFISAKMPIAYGWH	193
QY	239	WLLIISAV-PLLLFAVICFWLPESARYDVLL-SGNOEKAIATLKRIATENGAMPLOGLII	296
DB	184	SLLVIGGVLPPLLALVLMVWLPESARFLVVNRGTDKIRKTLSPIAQVVA--EAGSFVS	241
QY	297	SROE---DRCKMRDLTPTHRTWTLLLWFWSNATSYCYGLVLTTELFOAGDVCGLSSR	353
DB	242	PEOKAAVASRVFAV-FSGTYGLGTMLLWLYFMGLVIIVY---LLTSMPLTMLMDSGASHE	298
QY	354	KKAVERAKSLACEYLSEEDYMDLLTTLSEFPGLVLT---WIIDLRLGRKKTMALCFV--	408
DB	299	QAFIG-----ALFQGGVLSAVGVGWANDRYPNHKVIGIIFVLLA	338
QY	409	-IFSFCULLFICVGRN----VLTLLLFITAFISGGFOAAVYVTEPYPTATRALGLG	452
DB	339	GVFAY-----AVGOSLGNTITLATLVLIAGMCVNGSAQSAMPASAAREFYPTOGRTGVS	391
QY	463	TCSGMARVGAALIIPFFIAQVNMLESSVYLTLLAVYSGCCLL	500
DB	392	WMIGIGRGFGAI-----LCAWSGATLL	412

RESULT 6

PCAK_ACICA	PCAK_ACICA	STANDARD;	PRT;	457 AA.
ID	PCAK_ACICA	STANDARD;	PRT;	457 AA.
AC	O43975;			
DC	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	4-HYDROXYBENZONATE TRANSPORTER.			
DN	PCAK.			
OS	Acinetobacter calcoaceticus			
OC	Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;			
OC	Acinetobacter.			
OX	NCBI_TaxID=471;			
ON	[1]			
RC	SEQUENCE FROM N.A.			
RX	STRAIN=BD413 / ADP1;			
RX	MEDLINE=94341565; PubMed=8063101;			
RA	Kowalchuk G.A., Hartnett G.B., Benson A., Houghton J.E., Ngai K.-L.,			
RA	Ornston L.N.;			
RT	*Contrasting patterns of evolutionary divergence within the			
RT	Acinetobacter calcoaceticus pca operon.;			
CC	Gene 146:23-30(1994).			
CC	-1 SURCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE			

```
CC CC (POTENTIAL).
CC CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC DR EMBL; L05770; AAC37151.1; -.
CC DR InterPro; IPR003662; sub_trnsporttr.
CC DR Pfam; PF00083; sugar_tr; 1.
CC DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
CC DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
CC KW Transport; Transmembrane; Inner membrane.
CC FT TRANSMEM 35 55 POTENTIAL.
CC FT TRANSMEM 73 93 POTENTIAL.
CC FT TRANSMEM 102 122 POTENTIAL.
CC FT TRANSMEM 129 149 POTENTIAL.
CC FT TRANSMEM 169 189 POTENTIAL.
CC FT TRANSMEM 192 212 POTENTIAL.
CC FT TRANSMEM 275 295 POTENTIAL.
CC FT TRANSMEM 311 331 POTENTIAL.
CC FT TRANSMEM 339 359 POTENTIAL.
CC FT TRANSMEM 365 385 POTENTIAL.
CC FT TRANSMEM 401 421 POTENTIAL.
CC FT TRANSMEM 427 447 POTENTIAL.
CC SQ SEQUENCE 457 AA; 49277 MW; 4F5B5F77361A1567 CRC64;

Query Match 12.2%; Score 346.5; DB 1; Length 457;
Best Local Similarity 24.9%; Pred. NO. 1.le-17;
Matches 115; Conservative 92; Mismatches 197; Indels 57; Gaps 14;

QY 56 VPKEFANPTDDTFMVE---DAVEAIG---FGKFWKLSVLTGLAWMADAMEMWTLSTA 108
DB 1 MPKE-ANMASQDYATQRSLSQAQALINDAPLSYQWLAIIVCFLLVFDGDTAAMGRIA 59

QY 109 POLHCEWRLPSQVALLTSVYVGMSSSTLWGNISDQYGRKTGLKISVLWTLTYGILSA 168
DB 60 PALAQDWGVDRSQLGPNVNSAALGGMIIGALVSGPTADRFGRKIVLSMSLVFGGFTLACA 119

QY 169 FAPVYSWILVLRLGLVFGFVG-GVPOSVTLYAELFKPMKARAKILLIEFVWAGTVEVL 227
DB 120 YSTNLSLVIFRFLTGIGLGAAMPNATTLFSEYCPARISRLVTCMFCGYNLGMAIGFI 179

QY 228 AVFVMPSLGWRWLLLSA-VPLLFAVLFCFPLPESARYDVLGG-NOEKAIATLKRIA--- 282
DB 180 SSWLIPAFGWSHSLFLGGWAPLMLLVIFLFPESYRELIVKGNKTKVQILSRIPQK 239

QY 283 ----TENGAPMLPKLIIISQEDRGKMRDLTPHFRWTLTLLWTFWNSAASYGLVLT 338
DB 240 VQGVTEFHVP-----BEKVEAGTKKGVFGVSKYVKGTVLWTVYFMGLVMYI---LLT 292

QY 339 TELFQAGDVGCGTSSRRKKAWEAKSLACEVLSSEYMDLLWTLTSLFPGVLT---WIID 395
DB 293 SWLPTLMRETGASLERAA-----FLG-----GLFQFGVLSALFIGWAND 332

QY 396 RLGRKKTALCVIFSPCSLLLFICVGRN-----VLTLLFTIARFISGFOAAVYVYPE 450
DB 333 RENPNRIIA----GYLAAGIFAVIVGQSLNSPTLLALFILCAGIAGVANGAQSMPVLSAR 388

QY 451 VYPTATRALGLTCSGMARVAGLITPFIAQVMLESSVYLT 491
DB 389 FYPTOCATGVAMSGIGRFAGVAGVAGVILGNNSFTM 429

RESULT 7
OCN2_RAT
ID OCN2_RAT STANDARD; PRT; 557 AA.
AC Q70594; Q9QWL0;
```

```
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ORGANIC CATION/CARNITINE TRANSPORTER 2 (SOLUTE CARRIER FAMILY 22,
DE MEMBER 5) (HIGH-AFFINITY SODIUM-DEPENDENT CARNITINE COTRANSPORTER)
DE (UST2R) (CTI).
GN SLC22A5 OR OCTN2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=98200080; PubMed=9541011;
RA Schoenig E., Spitzberger F., Engelhardt M., Martel F., Oerding N.,
RA Gruendemann D.;
RT "Molecular cloning and characterization of two novel transport
RT proteins from rat kidney.";
RT FEBS Lett. 425:79-86(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Intestine;
RX MEDLINE=99011422; PubMed=9792817;
RA Sekine T., Kusuhaara H., Utsunomiya-Tate N., Tsuda M., Sugiyama Y.,
RA Kanai Y., Endou H.;
RT "Molecular cloning and characterization of high-affinity carnitine
RT transporter from rat intestine.";
RT Biochem. Biophys. Res. Commun. 251:586-591(1998).
RN [3]
RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=99384224; PubMed=10454528;
RA Wu X., Huang W., Prasad P.D., Seth P., Rajan D.P., Leibach F.H.,
RA Chen J., Conway S.J., Ganapathy V.;
RT "Functional characteristics and tissue distribution pattern of organic
RT cation transporter 2 (OCTN2), an organic cation/carnitine
RT transporter.";
RT J. Pharmacol. Exp. Ther. 290:1482-1492(1999).
CC -1- FUNCTION: SODIUM-ION DEPENDENT, HIGH AFFINITY CARNITINE
CC TRANSPORTER. ALSO TRANSPORTS ORGANIC CATIONS WITHOUT THE
CC INVOLVEMENT OF SODIUM. INVOLVED IN THE ACTIVE CELLULAR UPTAKE OF
CC CARNITINE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PROXIMAL AND DISTAL TUBULES
CC AND IN THE GLOMERULI IN THE KIDNEY, IN THE MYOCARDIUM, VALVES, AND
CC ARTERIOLES IN THE HEART, IN THE LABIRINTHINE LAYER OF THE
CC PLACENTA, AND IN THE CORTEX, HIPPOCAMPUS, AND CEREBELLUM IN THE
CC BRAIN.
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. ORGANIC
CC CATION SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC -----
CC EMBL; AJ001933; CAA05106.1; -.
CC EMBL; AB017260; BAA34399.1; -.
CC EMBL; AF110416; AAB54059.1; -.
CC InterPro; IPR003662; sub_trnsporttr.
CC Pfam; PF00083; sugar_tr; 1.
CC PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
CC Transport; Transmembrane; Glycoprotein.
CC TRANSMEM 21 41 POTENTIAL.
CC TRANSMEM 143 163 POTENTIAL.
CC TRANSMEM 173 193 POTENTIAL.
CC TRANSMEM 198 218 POTENTIAL.
CC TRANSMEM 233 253 POTENTIAL.
CC TRANSMEM 258 278 POTENTIAL.
CC TRANSMEM 342 362 POTENTIAL.
```



Query Match	11.7%	Score 332;	DB 1;	Length 537;
Best Local Similarity	25.1%	Prod. No. 1.5e-16;		
Matches 105; Conservative	80;	Mismatches 172;	Indels	62;
Gaps	11;			

  

QY	114	EWRL---PSMQVALLTSVVFVGMKSSSTLGMNIGTSDOYGRKTLKISVLWTLVYGLLSAFA	170
Db	131	EWDLVCKDDMKAPLTLTSLFPVGLMGSPITSGQJSDRFGKRNVLFLTMGMGTGFSFLOVFS	190
QY	171	PVYSWILVLRLGLVFGF-IGGVPOSVLTLYAEFLPMKARAKCIL-IEVFWAIGVVFVEVLA	228
Db	191	VNPFEMETVLVIVGMGQISNVAFAEVLCPFLSKSTPITFATCIVCEVAGCEMVLDEFA	250

QY 229 YVMPSLGWRWLLILSAVPLLLFAVLCFWLPESARYDVLSGNOEKAIATLKRIATENCAP 288  
 DB 251 YFIR---DWMLLLALTVPVGLGALMMWFIPESPRLWISQRIKEAEVVIKAKINGIV 307  
 QY 289 MPLGLKLIISROEDRGKMDL-----FTPHFRWTTLLLMFWFNSAFSYGL 334  
 DB 308 AP-----STIFDPSELQDLNSTKPOLHHYDLIRNIRVITIMSIIILMTISGVYGL 361  
 QY 335 VLLTTTELFOAGDVCYISRRKKAKEAKCSLACEYSEEDYMDLLWTTLSEFFCVLVTLMII 394  
 DB 362 SLDTNLH--GDI-----YVNCFLAAVEVPAYVLWALL 394  
 QY 395 DRLGKKTALCFVIFSCSLLLFICVGRNVL---TLLLFIAFISGGFOAAVYVYPE 450  
 DB 395 QYLPRYSISA--ALFLGCVLLFWQLVPSLSELFYSTALVWVGKGIITSAYSMVYVYAE 452  
 QY 451 VYPTATRALGTCGSMARVAGALITPFTIAQWMLSSVYTLAVYSGCCLLAALASCFPL 509  
 DB 453 LYPTVVRNMGVGVSTASRLGSIILSPYFVYLGAAYR-FLPYILMGSLTILTAITLFFP 510

RESULT 9  
 YIAJ\_BACSU  
 ID YIAJ\_BACSU STANDARD; PRT; 451 AA.  
 AC P37514;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YIAJ.  
 YIAJ.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96051385; PubMed=7584024;  
 RA Ogatawara N., Nakai S., Yoshikawa H.;  
 RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
 RT subtilis chromosome containing the replication origin.";  
 RL DNA Res. 1:1-14(1994).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D26185; BAA05214.1; -  
 CC EMBL: Z99124; CAB16121.1; -  
 CC Subtilist; BG10042; YIAJ.  
 CC InterPro: IP0003662; sub.transpor.  
 CC Pfam: PF00083; sugar\_tr\_1  
 CC PROSITE: PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
 CC PROSITE: PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
 CC KW Hypothetical protein; Sugar Transport; Transmembrane; Complete proteome.  
 FT TRANSMEM 30 50 POTENTIAL.  
 FT TRANSMEM 61 81 POTENTIAL.  
 FT TRANSMEM 99 119 POTENTIAL.  
 FT TRANSMEM 128 148 POTENTIAL.  
 FT TRANSMEM 159 179 POTENTIAL.  
 FT TRANSMEM 183 203 POTENTIAL.  
 FT TRANSMEM 271 291 POTENTIAL.  
 FT TRANSMEM 306 326 POTENTIAL.  
 FT TRANSMEM 345 365 POTENTIAL.  
 FT TRANSMEM 397 417 POTENTIAL.  
 FT TRANSMEM 422 442 POTENTIAL.

SQ SEQUENCE 451 AA: 49672 MW: 4A00D3891C4D7D09 CRC64;

Query Match 11.3%; Score 322; DB 1; Length 451;  
 Best Local Similarity 25.2%; Pred. No. 6.3e-16;  
 Matches 117; Conservative 71; Mismatches 204; Indels 72; Gaps 14;

QY 74 VEAIGFGKFWKLSVLTGLAWMADAMEMMILSIAPOLHCEWRPLPSWOVALLTSVVFVGM 133  
 DB 20 LDRPLISRVHFQVLTALGIVFFDLADLTLSNAPALIEHMGIPLSLTIANVTAASFLGM 79  
 QY 134 MSSSTLWGNISDOYGRKTGLKISLVLTLYGILSAFAPVYSMILVRLGVFGIGVGPQS 193  
 DB 80 FLGASLGLGRSLDRIGRKALNLVVFVSIASLCNAAWMDIPSLMTFRFLTGFVGAAMVI 139  
 QY 194 VTLV-AEELPMKARAKICILLIEVFWAIGTVFVAVLAVMP--SLGWRWLLILSAVPLLL 250  
 DB 140 TDSYLAEEFPFSVRGKYISFCAMIGLIGVPTNIVSAFVPLGSMGWRVLFVWGAAGLIY 199  
 QY 251 FAVLCFWLPESARYIVLVSGNOEKAIATLKRI---ATENGAMPPLG---KLIISROEDRG- 303  
 DB 200 FFFI-HRLEESPRWHENRGYAKADAILTRIEQVEKEKGLPLPAASQPKVSETVKNAGY 258  
 QY 304 ---KMRDLFTPHFRWTTLLLMFWFNSAFSYVGLVLTTLTTELFOAGDVCYISRRKKAVEA 359  
 DB 259 AGLLGRNL-----KITIVLSAVWIFETFGFYGFASVPSLLKXNGV----- 300  
 QY 360 KCLACEYLSEEDYMDLLTTLSEFP-----GVLTWLTIIDLRKKTALCFVIFSCSL 415  
 DB 301 -----TMENTLWYNVLHVGAPLGLALLGSMISERFORFWILAAASFLTAIAGL 348  
 QY 416 LLFICVGRNVLTLFLI-----ARAFISGGFOAAVYVYPTATRALGLTCGSM 467  
 DB 349 L-----YGMTFIPIMIIIVFGFIVITERVFTSN-----LYATSEPTPEYRSSGSLAYGL 400  
 QY 468 AR---VGALETPFIAQWMLSSVYTLAVYSGCCLLAALASCF 507  
 DB 401 GRFSNIFGSLLVGFIAVQLGISVFLFI---GGCWLACSLLLIF 441

RESULT 10  
 MUCK\_ACICA  
 ID MUCK\_ACICA STANDARD; PRT; 413 AA.  
 AC P94131;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE CIS,CIS-MUCONATE TRANSPORT PROTEIN.  
 GN MUCK.  
 OS Acinetobacter calcoaceticus.  
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
 OC Acinetobacter.  
 OX NCBI\_TaxID=471;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BD413 / ADP1;  
 RX MEDLINE=97440147; PubMed=92944455;  
 RA Williams P.A., Shaw L.E.;  
 RT "muck", a gene in Acinetobacter calcoaceticus ADP1 (BD413), encodes the  
 RT ability to grow on exogenous cis,cis-muconate as the sole carbon  
 RT source.";  
 RL J. Bacteriol. 179:5935-5942(1997).  
 CC -!- FUNCTION: PROBABLE UPTAKE OF MUCONATE.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (POTENTIAL).  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[1]  
SEQUENCE FROM N.A.  
STRAIN=BD413 / ADPL;  
MEDLINE=97440148; PubMed=9294456;  
Collier L.S., Nichols N.N., Neidle E.L.;  
"benk encodes a hydrophobic permease-like protein involved in benzoate  
degradation by Acinetobacter sp. strain ADPL";  
J. Bacteriol. 179:5943-5946(1997).  
CC -!- FUNCTION: PROBABLE UPTAKE OF BENZOATE.  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
(POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF009224; AAC46425.1; .  
DR InterPro: IPR003662; sub-transporter.  
DR Pfam; PF00083; sugar\_tr; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Transport; Transmembrane; Inner membrane.  
FT TRANSMEM 23 43 POTENTIAL.  
FT TRANSMEM 61 81 POTENTIAL.  
FT TRANSMEM 94 114 POTENTIAL.  
FT TRANSMEM 120 140 POTENTIAL.  
FT TRANSMEM 151 171 POTENTIAL.  
FT TRANSMEM 182 202 POTENTIAL.  
FT TRANSMEM 259 279 POTENTIAL.  
FT TRANSMEM 298 318 POTENTIAL.  
FT TRANSMEM 326 346 POTENTIAL.  
FT TRANSMEM 349 369 POTENTIAL.  
FT TRANSMEM 388 408 POTENTIAL.  
FT TRANSMEM 415 435 POTENTIAL.  
SQ SEQUENCE 466 AA; 51403 MW; D60C2E61D512F334 CRC64;

Query Match 10.4%; Score 295.5; DB 1; Length 466;  
Best Local Similarity 23.0%; Pred. No. 5e-14;  
Matches 104; Conservative 82; Mismatches 179; Indels 87; Gaps 12;  
  
QY 82 FOWKLSVLTGLAWADAMEMMILSLAPQLHCEWRLPSSQVALLTSVVFVGMMSSTLWG 141  
DB 19 FHWVILSTLIIFDGYDLVIYGVVALPMLKWEAIDPVTAGTIGSIALFCMMFCALLFG 78  
  
QY 142 NISDQ-----YGRKTGLKISVLTLYYIGLSAFAPVYVSMILVLRGLVGFGGV-PQSVT 195  
DB 79 TIADKLEHLGVSRKKVIAVCIIILFSLCTVLCGSETTTQFSIFRFLAGVIGGVMPNVIA 138  
  
QY 196 LYAEFLPMKARAKICILLIEVFAIGTVEVVLAVFVMSLGRWLLIILSAVPLLLFAVLC 255  
DB 139 LVSEYAPKKFKSFVTLMFSGYAGMTAAFLGSLVPLYGKIMFMFIAGIPLVLLPLM 198  
  
QY 256 FWLPESARYDVLSGNQEKAIATLKRIAT-----ENGAPMPLGLKI 295  
DB 199 EVLPESIDYLV-----RKKDETFRIMTKMPSYQYQPDHVFVLNSSNQQAQAPV-KMI 253  
  
QY 296 ISROEDRGKMRDLTPHFRWTLTLWFTWFSNAPSYGLVLLTTLFELFOAGDVCGISSRKK 355  
DB 254 F--QEORA-----FSTMMFWSICFIMTLINVYAL-----GNWLPKL 286  
  
QY 356 AVEAKCSLACEYLSLSEEDYMDLLTTLSEFPGVLVTLWIDRLGRK-----KTMALC 406  
DB 287 MIEAGYNLS-----KSLIFLSNVGMIGSLGGLYLDNRVNVKFTWGLLLLCALISLS 340  
  
QY 407 FVIFSFCSLLIFCVGRNVLTLLFIARAFISGGQAAVYVTPYVPTATRALGTCGSG 466  
DB 341 LLSFQFSSVILYI-----LIACAGAAISGAIQIMLLAYMAKFPAPNVRSTGIGWGLG 391  
  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
EMBL; U87258; AAC27117.1; .  
DR InterPro: IPR003662; sub-transporter.  
DR Pfam; PF00083; sugar\_tr; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Transport; Transmembrane; Inner membrane.  
FT TRANSMEM 17 37 POTENTIAL.  
FT TRANSMEM 54 74 POTENTIAL.  
FT TRANSMEM 86 106 POTENTIAL.  
FT TRANSMEM 113 133 POTENTIAL.  
FT TRANSMEM 146 166 POTENTIAL.  
FT TRANSMEM 172 192 POTENTIAL.  
FT TRANSMEM 229 249 POTENTIAL.  
FT TRANSMEM 267 287 POTENTIAL.  
FT TRANSMEM 294 314 POTENTIAL.  
FT TRANSMEM 319 339 POTENTIAL.  
FT TRANSMEM 362 382 POTENTIAL.  
FT TRANSMEM 383 403 POTENTIAL.  
SQ SEQUENCE 413 AA; 45245 MW; 2D88CE31C4C5CC65 CRC64;  
  
Query Match 10.4%; Score 295.5; DB 1; Length 413;  
Best Local Similarity 24.9%; Pred. No. 4.4e-14;  
Matches 111; Conservative 80; Mismatches 179; Indels 75; Gaps 18;  
  
QY 80 GKFOVKLSVLTG-LAWMADAMEMMILSLAPQLHCEWRLPSSQVALLTSVVFVGMMSST 138  
DB 11 GSHTWKIAFLFAFLALVDGADMLLSYLSNSIKAEFNLSVTEAGMLGSETFLAGMAIGGI 70  
  
QY 139 LWMNISDQYGRKTKLSVLTWLYYIGLS-AFAPVYSWIL--VLRGLVGFVGIGVQSV- 194  
DB 71 FCGWACDRFGR--VRIWVISILTSILTCGLGTQSFIOFQVLRFPFASLGLSLYIACN 127  
  
QY 195 TLYAEFLPMKARAKICILLIEVFAIGTVEVVLAVFVMSLGRWLLIILSAVPLLLFAVL 254  
DB 128 TLMAEYVPTKYRTVLTGLOAGTGVIVATLLAGLIPDHGRVLFYVAIIPVLMVLM 187  
  
QY 255 CFWLPESARYDVLSGNQEKAIATLKRIATENGAPMPLGLKLIISROEDRGKMRDLTPHFR 314  
DB 188 HFFVPEPAAM-----QOSRLAPSKQFETVKTSAF---KLIFQDKRNRN----- 227  
  
QY 315 WTTLLLFNIFNSNA---FSYIGLVLLTTLFQAGDVCYGISSRKKAVEAKCSLACEYLS 370  
DB 228 -----MFLMALTAGFLQFGYGVNNMMPSVLES-----ELGMKEKEM 265  
  
QY 371 EDYMDLLTTLSEFPGLVLTWIDRLGRKKTALCEV-IFSFCSLLLFITCVGRNVLTLL 429  
DB 266 TAYMVGTYTAM--ILGKILAGFMADKLGRRFTYAFGAIGTAIFLPLIVFYNSPDNIIYLL 323  
  
QY 430 LFIARAF SG---GFOAAVYVTPVYPTATRALGLGTCSGMARVGAITP----FIAQVM 482  
DB 324 VIF--GFLYGIPIGVNATYM--TESFTAIRGTAIGAYNVGRGLGAIAIAPATIGFLAS-- 377  
  
QY 483 LESSVYLT-----AVYSGCCLLAL 503  
DB 378 -GSGTGLGFVVMGAAYFICGVIPAL 401  
  
RESULT 11  
BENK\_ACICA STANDARD; PRT; 466 AA.  
ID BENK\_ACICA AC 030513;  
DT 20-AUG-2001 (Rel. 40, Created)  
DT 20-AUG-2001 (Rel. 40, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE BENZOATE TRANSPORT PROTEIN.  
GN BENK.  
OS Acinetobacter calcoaceticus.  
OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;  
OC Acinetobacter.  
OX NCBI\_TaxID=471;

	7	7	F -> L (IN REF. 1).
CONFLICT	12	16	FSIIH -> CLIHV (IN REF. 1).
FT CONFLICT	30	30	D -> Y (IN REF. 1).
FT CONFLICT	115	115	T -> F (IN REF. 1).
FT CONFLICT	132	132	T -> N (IN REF. 1).
FT CONFLICT	240	240	R -> C (IN REF. 1).
FT CONFLICT	304	304	N -> S (IN REF. 1).
FT CONFLICT	395	395	N -> D (IN REF. 1).
FT SEQUENCE	443 AA;	48666 MW;	AER484DBB109DE05 CRC64;
Query Match 10.1%; Score 286.5; DB 1; Length 443;			
Best Local Similarity 22.9%; Pred. No. 2.1e-13;			
Matches 103; Conservative 92; Mismatches 216; Indels 39; Gaps 15;			
QY	92 LAW-----MADAMEMILSLAPOLHCEWRLPSPQVALLTSVVVGVMMSSSTLWGNISDQ 146		
Db	: : : :     : : : :     : : : :     : : : :     : : : :     : : : :		
Db	20 LLWGSGPFELDGVLVMIGVALGQLPALKDWDWICLLGAGTLAGLVGTSLFGVISDK 79		
QY	147 YGRKTKGLSVLTLYYLGYLSAFPVYSMTLVLRGLVGFGIGG-VPOSVTLYAEFLPMKA 205		
Db	: : : :     : : : :     : : : :     : : : :     : : : :     : : : :		
Db	80 VGRKKMFLLIDIIAIGVISVATMFVSSPVELLVWRVLIGIVIGADYPDATSMITEFSSTRQ 139		
QY	206 RARCIILLIEFWAIGTVFEVYLAVFVMP-SLGRMRWLILISAVPELLFAVLFCWLPESARY 264		
Db	: : : :     : : : :     : : : :     : : : :     : : : :     : : : :		
Db	140 RAFSIFIAAMWVVGATCADLVGYLWYDVGGNRWMLGSAAIPCLLILJGRFELPESP RW 199		
QY	265 DVLSNGOEKAIATLKRIATENGAPMLPGKLIIISROEDRGKMRDLFT-PHRFTWTLLLWF1 323		
Db	: : : :     : : : :     : : : :     : : : :     : : : :     : : : :		
Db	200 LLRKGVRVKEEEMIKLF---GEVPADF-----EQOQTRFDLFRNRRHPFF-VLFVAAL 251		
QY	324 WFSNAFSYVGLVLTITTELFQAGVDVCGISSRKKAVEAKCSLACEYLSEEDYMLLTWTLE 383		
Db	: : : :     : : : :     : : : :     : : : :     : : : :     : : : :		
Db	252 WTCQVIPMAFYIYFGPIV--GLLGIVGCKNAALGNVVIS-----LFFMLGC 296		
QY	384 FPOGLVTLWIIDLGRKKTWALCFVFPSFCSLLIFTCVGRNV-LTLLLFIARAFISGG-F 441		
Db	: : : :     : : : :     : : : :     : : : :     : : : :     : : : :		
Db	297 IPPN---LW-LNTAGRPLPLIGSFAMMTLALAVGLLIPDMGIMLVNMAFAVAYAFSGPG 352		
QY	442 QAAVYVTPVPTATRALGLTGCGMARVGALITPPTIAQVMLES-SVYLTLAVYSGCCLL 500		
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :		
Db	353 NLQWLPNELFPPTDIRASAVGVMSLSRICTIVSTWALPIFINNYGISINTMLMGAGISLF 412		
QY	501 AALASCFLPETKGGGLQESSHRE-WGOEM 529		
Db	: : : :     : : : :     : : : :     : : : :     : : : :     : : : :		
Db	413 GLLIISVAFAPETRGMSLAQTSNMRTIRGQM 442		
RESULT 13			
YLX5_CAEBL STANDARD; PRT; 751 AA.			
ID	YLX5_CAEBL	STANDARD;	PRT; 751 AA.
AC	P46501;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-NOV-1995 (Rel. 32, Last annotation update)		
DE	HYPOTHETICAL 84.8 KDA PROTEIN F23F12.5 IN CHROMOSOME III.		
GN	F23F12.5.		
OC	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;		
OX	Rhabditiidae; Pelodierinae; Caenorhabditis.		
NCBI_Taxid=6239;			
[1]			
SEQUENCE FROM N.A.			
RP STRAIN=BRISTOL N2.			
Du 2.;			
RA			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and/or commercial		

entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U12965; AAA20607.1; --  
CC WormPep; F23F12.5; CE01252.  
DR InterPro; IPR002184; Strb.  
DR Pfam; PF02175; Strb; 1.  
DR Pfam; PF00083; sugar.tr; 1.  
KW Hypothetical protein; Transmembrane.  
FT TRANSMEM 24 44 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT TRANSMEM 102 122 POTENTIAL.  
FT TRANSMEM 137 157 POTENTIAL.  
FT TRANSMEM 200 220 POTENTIAL.  
FT TRANSMEM 235 255 POTENTIAL.  
FT TRANSMEM 320 340 POTENTIAL.  
FT TRANSMEM 348 368 POTENTIAL.  
FT TRANSMEM 379 399 POTENTIAL.  
FT TRANSMEM 410 430 POTENTIAL.  
FT TRANSMEM 432 452 POTENTIAL.  
FT TRANSMEM 515 535 POTENTIAL.  
FT TRANSMEM 547 567 POTENTIAL.  
FT TRANSMEM 583 603 POTENTIAL.  
FT TRANSMEM 614 634 POTENTIAL.  
FT TRANSMEM 678 698 POTENTIAL.  
FT SEQUENCE 751 AA; 84832 MW; A6C4F43540295EFC CRC64;

Query Match 10.0%; Score 284; DB 1; Length 751;  
Best Local Similarity 25.1%; Pred. No. 5,3e-13;  
Matches 111; Conservative 82; Mismatches 149; Indels 100; Gaps 24;

QY 119 SNOVALLTSVVFVGMSSSTLMGNIHQYGRKT----GLKISVLWPLYGILSAFAPVYS 174  
DB 316 AYDAAVATQITGVIGALITYGHLGHDFGRKPVSPFGISVGIL----PCVASGFGAPSW 371  
QY 175 WILVLCGLVCGIGGVPQSVTLVLAEPKPKAKCILLIEVF-WAIGTVFVVAVFVMP 233  
DB 372 VFAAFRPIVGTSTASI--LIVFYAYILEFIEPQRFVLSFFNWGYARLV-FTLACFICG 428  
QY 234 SLGWR-----WLLLSAVPLLFVAVLCFWLPESARYDVLSCNOEKAIAATLKRATFENGA 287  
DB 429 Y--WRSAIAATSLSLPILPVLLI-----LPSPKWNFKRFRDARAARVAVLMSGI 480  
QY 288 P-----MPLGKLLISQEDRGKRDLPFPHRT----TLLLWFIWFSNAFVYGLV 335  
DB 481 PYVNDEQDSIESEKLEEKSTKIYTKMDLFT----SWTIAYRTIVVGLSFWSTLSAFGSD 537  
QY 336 LLTTELFQAGDVCGIISRRKKAVERKSLACEYLSEEDYMDLLTTLSEFPGLVLTWIID 395  
DB 538 L-----NSGNLAGNF-----YLSQ-----FVSGAVTAFAKIFV--FLLD 569  
QY 396 -----RLGRKKYTMALCFVIFSCSLLLFICV-----GRVLTLL--FIARAF 436  
DB 570 TYVPSEDRRLHQYQIAM---ILCYVIMILPESDQSGQSRDLAIHINIIGVSF 626  
QY 437 ISGFGQAAVYVTEVYPTATRALGLGTCGSMARVGALITPFAQVLMLESSVYLTAVYSG 496  
DB 627 IEITWDACYLVAVECPFKTIRTIIGTCSSLARTGALLAPQAYL---SDIYRP-APYAV 682  
QY 497 CC---LALALASC-FPIETKG 514  
DB 683 VCSIGTISLLISCVFLP-DPKG 703

RESULT 14  
YGS\_ECOLI STANDARD; PRT; 445 AA.  
AC Q46909;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN YGS.  
GN YGS OR B2771.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT The complete genome sequence of Escherichia coli K-12.\*;  
RL Science 277:1453-1474(1997).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
(POTENTIAL).  
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U29579; AAA69281.1; ALT\_INIT.  
DR EMBL; AE000360; AAC75813.1; ALT\_INIT.  
DR EcoGene; EG13126; YGS.  
DR InterPro; IPR003662; sub.transpstr.  
DR Pfam; PF00083; sugar.tr; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;  
KW Complete proteome.  
FT TRANSMEM 23 43 POTENTIAL.  
FT TRANSMEM 57 77 POTENTIAL.  
FT TRANSMEM 86 106 POTENTIAL.  
FT TRANSMEM 115 135 POTENTIAL.  
FT TRANSMEM 143 163 POTENTIAL.  
FT TRANSMEM 176 196 POTENTIAL.  
FT TRANSMEM 254 274 POTENTIAL.  
FT TRANSMEM 287 307 POTENTIAL.  
FT TRANSMEM 312 332 POTENTIAL.  
FT TRANSMEM 338 358 POTENTIAL.  
FT TRANSMEM 370 390 POTENTIAL.  
FT TRANSMEM 401 421 POTENTIAL.  
FT SEQUENCE 445 AA; 48234 MW; B59E452721B15774 CRC64;

Query Match 9.9%; Score 282.5; DB 1; Length 445;  
Best Local Similarity 24.0%; Pred. No. 4e-13;  
Matches 114; Conservative 82; Mismatches 212; Indels 67; Gaps 15;

QY 74 VEAIGKTFQWKLVSUTGLAWMADAMMTLSILAPOLHCENRPLSQVALLTSVVFVGM 133  
DB 8 MDOLPLNFRHCRIAALTFGAHLTDGVVLGVYVAILQITPAMQITPFMAQMGISALLGL 67  
QY 134 MSSSTLGNISDOYGRKTGLKISVLWTLVYVILSAFAPVYSWILVLRGLVGFVGVPQS 193  
DB 68 FLGSLVLGNISDHIGRQKIFTFSLITLASFLQFFATTPHEHLIGLILIGLGG-DYS 126  
QY 194 V--TLVAEFLPMKARAKCILLIEVFWAIGTVFVAVFVMPSL-----GHRWL 240  
DB 127 VGHITLAEFSPPRRHGIILGAFSVVMTVG-----YVLASTAGHHFISENPEARWL 177  
QY 241 LILSAVPLLLFAVLCFWLPESARYDVLSCNOEKAIAATLKRATENGAPMLGLIISRQE 300  
DB 178 LASAALPALLITLLRWGTPEPSRWLRQGRFAEHAHVHYF---GPHVLLGDEVVTATH 234  
QY 301 DRGKMRDLTFPHFRWTLTLLWFIFWFSNAFVYVGLVLTTLTTELFQAGDVCGIISRRKKAVERK 360

Db 235 KH--IKTFSSRY-----WRTAFNSVFFVCLVIPWF-----VIYTW 269  
QY 361 CSLACEYLSEEDYM--DLLWTTLSEPPGVLTWIIIDRLGRKKTALCFVIFSCSLLLF 418  
Db 270 LPTIAOTIGLEDALMALNALL-IVGALLGLVTHLHAHRFLUGSULLA-ATLVVM 327  
QY 419 ICV-GRNVLTLLIFTA-RAFISGGFOAAVYVTPETATRALGLGTCGSMARVGAII-T 475  
Db 328 ACLPSGSSLTLLLVLFSTISAVNSLVGLIPAESPTDIRSIGVGFTATMSRLGRAVST 387  
QY 476 PFTAQVWLESSVYLTAVYSGCCLLAALASCFLPIETKG-----GGLQSS 521  
Db 388 GLLPWLMQMGQVTLILLATVLLGVFVVTWLPAPETKALPLVAAGNVGGANERS 442

RESULT 15  
YB04\_HAEIN  
ID YB04\_HAEIN STANDARD; PRT; 407 AA.  
AC P71369;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN H1104.  
GN H1104.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID: 127;  
RN [1]  
RC SEQUENCE FROM N.A.  
RP STRAIN-RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kervavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-F., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
CC (POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U32790; AAC22759.1; -  
DR TIGR: H1104; -  
DR InterPro: IPR003662; sub.transprtr.  
DR Pfam: PF00083; sugar.tr; 1  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; FALSE\_NEG.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;  
KW Complete proteome.  
FT TRANSMEM 17 37 POTENTIAL.  
FT TRANSMEM 49 69 POTENTIAL.  
FT TRANSMEM 78 98 POTENTIAL.  
FT TRANSMEM 108 128 POTENTIAL.  
FT TRANSMEM 139 159 POTENTIAL.  
FT TRANSMEM 161 181 POTENTIAL.  
FT TRANSMEM 225 245 POTENTIAL.  
FT TRANSMEM 262 282 POTENTIAL.

FT TRANSMEM 289 309 POTENTIAL.  
FT TRANSMEM 313 333 POTENTIAL.  
FT TRANSMEM 358 378 POTENTIAL.  
FT TRANSMEM 379 399 POTENTIAL.  
SQ SEQUENCE 407 AA; 43681 MW; 6185886CB37C4A9 CRC64;  
  
Query Match 9.9%; Score 280.5; DB 1; Length 407;  
Best Local Similarity 23.0%; Pred. No. 5e-13;  
Matches 107; Conservative 77; Mismatches 185; Indels 97; Gaps 16;  
  
QY 82 FQWKLSTLTGLAMMADAMEMWLSILAPQLHCEWRLPSQVALLTSVVVFGMMSSTLWG 141  
Db 8 YGKALIGSAVGYGMDGFDLLILGFMLSAISADLNLTPAOGCSLVTTWTLIGAVFGGILFG 67  
  
QY 142 NISI QYGRKTKLSVLWLTLYYGILSAFAPVYSWILVLRGLVGFICGG-VQPSVTLVYAEF 200  
Db 68 ALSDKYGRVRVLTWTLILFAVFTGLCAIQGYWDLIIYRTIAGIGLGGEPGIGMALAAEA 127  
  
QY 201 LPMKARAKCILLIEVFVAIGTVFEVLAVFVMPSLGWRWLLILSAVPLILLFAVLCFWLPE 260  
Db 128 WPAHRAKAASYVALGHQGVGLGAALLTPILLPHIGRMGLVG-----IFPAFVAMELR 182  
  
QY 261 SARYDVLSGNOEKAIATLKRIATENGAPMPGLKLIISQEDRGKMRDLFTPHKRWTTLL- 319  
Db 183 SHLH-----EPEIFTQKQTA-----LSTQSS-----FTDKLRSFOLLI 215  
  
QY 320 -----LWFIWFSNA--FSYVGLVL-----LTTELFAQGDVCGISSRKKAVEAKCSL 363  
Db 216 KDKATSKISIGIVLTQVQNFYGGIWIWLPNLSKQL-----GFSLTSG----- 261  
  
QY 364 ACEYLSSEEDYMDLLWTTLTSEFPGLVTLWIT-----DRLGRKKTALCFVIFSCSLLFI 419  
Db 262 -----LWTAVT-VCGMMAGIWFQGLADRIGRKPS-----FLLFOLGAVISIV 303  
  
QY 420 CVGRNVLTLLIFTARAFIS-----GGFQAAVYVTEVYPTATRALCLGTCGSMAR-VG 471  
Db 304 VYSQLTDPDMLLAGAFGLGMFVNGMLGGYCALMA---EAYPTARATAQNVLFNTGRAVG 360  
  
QY 472 ALITPPTAQVWLESSVYLTAVYSGCCLLAALASCFLPIETKGGGL 517  
Db 361 GFGPVVVGSVVLAY:FOATALALLAIIVVIDMLATIFLIPELKGKAL 406

Search completed: March 13, 2002, 12:42:52  
Job time: 158 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2002, 12:40:14 ; Search time 43.34 Seconds  
(without alignments)  
1849.498 Million cell updates/sec

Title: US-09-911-667a-2  
Perfect score: 2846  
Sequence: 1 MEEDLQRLQPLPVVKFRRTG.....MVGRGMHGAGVTRNSGSQ 548

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL\_17:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2749	96.6	548	11 Q92217	Q92217 rattus norv
2	1129.5	39.7	497	5 Q9WLD4	Q9WLD4 drosophila
3	863	30.3	470	10 Q9LSH7	Q9LSH7 arabidopsis
4	855	30.0	500	10 Q9M7W2	Q9M7W2 arabidopsis
5	734	25.8	144	4 Q9NPW5	Q9NPW5 homo sapien
6	644.5	22.6	454	2 Q9RYN5	Q9RYN5 deinococcus
7	509.5	17.9	422	2 Q9X1T9	Q9X1T9 thermotoga
8	466.5	16.4	455	2 Q9HYD9	Q9HYD9 pseudomonas
9	446	15.7	709	5 Q9W3W9	Q9W3W9 drosophila
10	436	15.3	593	11 P70485	P70485 rattus norv
11	435	15.3	555	11 Q9R0W2	Q9R0W2 rattus norv
12	435	15.3	593	11 P97558	P97558 rattus norv
13	432	15.2	535	11 Q63314	Q63314 rattus norv
14	429.5	15.1	556	11 Q63089	Q63089 rattus norv
15	429	15.1	554	6 Q02713	Q02713 sus scrofa
16	427	15.0	554	4 Q15395	Q15395 homo sapien
17	425	14.9	554	4 Q15245	Q15245 homo sapien
18	425	14.9	556	11 Q08966	Q08966 mus musculu
19	423.5	14.9	556	11 Q9R1Q4	Q9R1Q4 mus musculu

20	423	14.9	553	11 Q70577	Q70577 mus musculu
21	423	14.9	553	4 Q15244	Q15244 homo sapien
22	420	14.8	727	11 Q92216	Q92216 rattus norv
23	417	14.7	554	4 Q9NOD4	Q9NOD4 homo sapien
24	409.5	14.4	683	4 Q94840	Q94840 homo sapien
25	409	14.4	554	6 Q77504	Q77504 oryctolagus
26	409	14.4	742	11 Q9JIS5	Q9JIS5 mus musculu
27	403.5	14.2	742	4 Q94841	Q94841 homo sapien
28	403	14.2	742	6 Q29397	Q29397 bos taurus
29	402.5	14.1	683	11 Q63564	Q63564 rattus norv
30	398.5	14.0	562	13 Q57379	Q57379 pseudopleur
31	398	14.0	537	11 Q88909	Q88909 mus musculu
32	396.5	13.9	536	11 Q9R1U7	Q9R1U7 rattus norv
33	393.5	13.9	724	13 Q90406	Q90406 discopryge o
34	393.5	13.8	430	11 Q35882	Q35882 rattus norv
35	393	13.7	607	4 Q9BV29	Q9BV29 homo sapien
36	389.5	13.7	443	2 Q9L7Y1	Q9L7Y1 pseudomonas
37	386.5	13.6	551	4 Q9Y226	Q9Y226 homo sapien
38	379	13.3	551	11 Q88446	Q88446 rattus norv
39	376	13.2	575	5 Q9BKR9	Q9BKR9 caenorhabdi
40	375	13.2	551	11 Q9WTW5	Q9WTW5 mus musculu
41	375	13.2	551	11 Q99JF0	Q99JF0 mus musculu
42	374.5	13.2	563	4 Q95742	Q95742 homo sapien
43	371	13.0	539	4 Q9H2W5	Q9H2W5 homo sapien
44	368	12.9	526	10 Q9LHQ6	Q9LHQ6 arabidopsis
45	367.5	12.9	548	5 Q9VCA2	Q9VCA2 drosophila

ALIGNMENTS

RESULT 1  
ID Q92217 PRELIMINARY; PRT: 548 AA.  
AC Q92217;  
DT 01-MAY-1999 (Tremblrel. 10, Created)  
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE SV2 RELATED PROTEIN.  
GN SVOP.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=99019745; PubMed=9801366;  
RA Janz R., Hofmann K., Sudhof T.C.;  
RT "SVOP, an evolutionarily conserved synaptic vesicle protein, suggests  
RT novel transport functions of synaptic vesicles.";  
RL J. Neurosci. 18:9269-9281(1998).  
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -|- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL: AF060173; AAC78627.1; ..  
DR InterPro: IPR003662; sub-transporter.  
DR Pfam: PF00083; sugar\_cr; 1.  
KW Transmembrane.  
SQ SEQUENCE 548 AA: 60804 MW: 19AD8475B7579496 CRC64;

Query Match 96.6%; Score 2749; DB 11; Length 548;  
Best Local Similarity 96.0%; Pred. No. 1.1e-178;  
Matches 526; Conservative 10; Mismatches 12; Indels 0; Gaps 0;

QY 1 MEEDLQRLQPLPVVKFRRTGESARSDDDTASGHEVQIEGVHVGLAEVLDDGAAVPKEF 60

Db 1 MEEDLQRLQPLPVVKFRRTGESARSDDDTASGHEVQIEGVHVGLAEVLDDGAAVPKEF 60

QY 61 ANPTDDTFWEDAVEAIGFGKQFQWKLVSITGLANMADAMEMMILSLAPQLHCEWRLPSW 120

Db 61 ANPTDDTFWEDAVEAIGFGKQFQWKLVSITGLANMADAMEMMILSLAPQLHCEWRLPSW 120





OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA:  
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-COLUMBIA:  
 RX MEDLINE=20277480; PubMed=10819329;  
 RA Nakamura Y.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
 features of the regions of 4,504,864 bp covered by sixty P1 and TAC  
 clones.";  
 RL DNA Res. 7:131-135(2000).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
 EMBL; AB026645; BAB02515.1; -;  
 DR InterPro; IPR003662; sub.transporter.  
 DR Pfam; PF00083; sugar tr; 1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
 KW Transmembrane.  
 SQ SEQUENCE 470 AA; 51457 MW; 1062413BA354616D CRC64;

Query Match 30.3%; Score 863; DB 10; Length 470;  
 Best Local Similarity 38.8%; Pred. No. 7-2e-51;  
 Matches 180; Conservative 92; Mismatches 168; Indels 24; Gaps 6;  
 QY 68 FMVEDAVEAIGFGKFWKLSVLTGLAWMADAMEMMILSLAPQLHCEWRLPSWQVALLTS 127  
 DB 8 FTVDEALVANGFGKFGQIYVLAYACMGWVAEAMEMMILSFVGPVQSLNLSARQESLITS 67  
 QY 128 VVFGVMSSTLWGNISDQYGRKTLKISVLWTLTYGILSAFAPVYSWILVRLGLVGEI 187  
 DB 68 VVFGMLIGAYSWGIVSDKHGRKGFITAVVTFVAGFLSAFSPNYMWLIILCLVGLGL 127  
 QY 128 VVFGVMSSTLWGNISDQYGRKTLKISVLWTLTYGILSAFAPVYSWILVRLGLVGEI 187  
 DB 68 VVFGMLIGAYSWGIVSDKHGRKGFITAVVTFVAGFLSAFSPNYMWLIILCLVGLGL 127  
 QY 188 GGVPSQVTLTYAEFLPMKARAKCILLIEVFWAIGTVFVFLVAVFVMPSLGWRWLLILSAVP 247  
 DB 128 GGGPVLASWYLEFIPAPSRGTWVVFSAFTVGTIFEASLAWLMPRLGWRWLLAFSSVP 187  
 QY 248 LLLFAVLCFWLPESARDVLSGNOEKAIALTKRIATENGAPMLGK-----IISROE 300  
 DB 188 SLLLLFYRWTSSEPRYLILQGRKAEALALEKIARMKTKQLPGLVLSSELELEENKE 247  
 QY 301 DRGMRDLFTPHFRWTTLLWFWFSAFVYGLVLTTELFOAGDVCYGSRRKKAVEAK 360  
 DB 248 PGFSLALLSPTLMKRILLWVFFGNFAYYGVVLTTELNNSHNRCYPTKEQ----- 301  
 QY 361 CSLACEYLSSEEDYMDLWTLSEPPGVLTWLIIDRLGRKTKMALCFVIFSCSLLLFIC 420  
 DB 302 ---LRNSNDVNYRDVFIAFSGFAGPGLLSAAMVDRLGRKASMA--SMLETCIFLLPLL 355  
 QY 421 VGRN--VLTLLLFARAFISGGFOAAVYVPEVPTATRALGLCTCGSMARVGLITPFI 478  
 DB 356 SHQSPFITVTLFGGRICISAAFTVWYIYAPETVAVRTGTVGSGVSGRIGILCLPV 415  
 QY 479 AQVMESSVYLTAV--YSGCCLLAALASCFLPIETKGGGLQES 520  
 DB 416 A-VGLVHCCHOTIAVLFEVVLVSGICVCLFPFETSGRDLTDS 458

RESULT 4  
 Q9M7W2 PRELIMINARY; PRT; 500 AA.  
 ID Q9M7W2;  
 AC Q9M7W2;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE PUTATIVE TRANSPORTER.  
 GN MGH6.16.

OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,  
 Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,  
 Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
 RT "Arabidopsis thaliana chromosome III P1 MGH6 genomic sequence."  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
 DR EMBL; AC024128; AAF35954.1; -;  
 DR InterPro; IPR003662; sub.transporter.  
 DR Pfam; PF00083; sugar tr; 1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
 KW Transmembrane.  
 SQ SEQUENCE 500 AA; 54573 MW; EC459EB09EF581C5 CRC64;

Query Match 30.0%; Score 855; DB 10; Length 500;  
 Best Local Similarity 36.4%; Pred. No. 2.7e-50;  
 Matches 180; Conservative 95; Mismatches 165; Indels 54; Gaps 7;  
 QY 68 FMVEDAVEAIGFGKFWKLSVLTGLAWMADAMEMMILSLAPQLHCEWRLPSWQVALLTS 127  
 DB 8 FTVDEALVANGFGKFGQIYVLAYACMGWVAEAMEMMILSFVGPVQSLNLSARQESLITS 67  
 QY 128 VVFGVMSSTLWGNISDQYGRKTLKISVLWTLTYGILSAFAPVYSWILVRLGLVGEI 187  
 DB 68 VVFGMLIGAYSWGIVSDKHGRKGFITAVVTFVAGFLSAFSPNYMWLIILCLVGLGL 127  
 QY 188 GGVPSQVTLTYAEFLPMKARAKCILLIEVFWAIGTVFVFLVAVFVMPSLGWRWLLILSAVP 247  
 DB 128 GGGPVLASWYLEFIPAPSRGTWVVFSAFTVGTIFEASLAWLMPRLGWRWLLAFSSVP 187  
 QY 248 LLLFAVLCFWLPESARDVLSGNOEKAIALTKRIATENGAPMLG----- 292  
 DB 188 SLLLLFYRWTSSEPRYLILQGRKAEALALEKIARMKTKQLPGLVLSSELELEENKN 247  
 QY 293 -----KLIISROEDRG--KMRDLFTPHFRWTTLLWFWFSAFV 330  
 DB 248 IPTENTHLLKAGESGEAVAVSKIIVLKADKPGFSLALLSPTLMKRILLWVFFGNFA 307  
 QY 331 YYGVLVLTTELFOAGDVCYGSRRKKAVEAKCSLACEYLSSEEDYMDLWTLSEPPGVLT 390  
 DB 308 YGVVLTTELNNSHNRCYPTKEQ-----LRNSNDVNYRDVFIAFSGFAGPGLIS 357  
 QY 391 LWIIDRLGRKTKMALCFVIFSCSLLLFICVGRN--VLTLLLFARAFISGGFOAAVYV 448  
 DB 358 AAMVDRLGRKASMA--SMLETCIFLLPLLSHQSPFITVTLFGGRICISAAFTVYIYA 415  
 QY 449 PEVPTATRALGLCTCGSMARVGLITPFI AQVMESSVYLTAV--YSGCCLLAALASC 506  
 DB 416 PEIYPTAVRTTGVGSGVSGRIGGILCLPLA-VGLVHCCHOTIAVLFEVVLVSGICVC 474  
 QY 507 FLPIETKGGGLQES 520  
 DB 475 LFPFETSGRDLTDS 488

RESULT 5  
 Q9NPW5 PRELIMINARY; PRT; 144 AA.  
 ID Q9NPW5;  
 AC Q9NPW5;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
 DE HYPOTHETICAL 15.4 KDA PROTEIN (FRAGMENT).  
 GN DKFZP761H039.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=AMYGDLA;  
 RA Blum H., Bauersachs S., Meves H.W., Weill B., Wiemann S.;  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 KW EMBL; AL359592; CAB94878.1; -;  
 DR Hypothetical protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 144 AA; 15387 MW; 769A51A9564FFCD CRC64;

Query Match 25.8%; Score 734; DB 4; Length 144;  
 Best Local Similarity 99.3%; Pred. No. 1e-42;  
 Matches 143; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 405 LCFVIFSCSLLLCVGRNVLTLLFTARAFISGGFOAAAYVYTPVPTATRALGLGTC 464  
 DB 1 LCFVIFSCSLLLCVGRNVLTLLFTARAFISGGFOAAAYVYTPVPTATRALGLGTC 60  
 QY 465 SGMARVALITPPIAQMVLSSVYTLAVYSGCCLLAALASCFLEPIETKGGGLQESSHRE 524  
 DB 61 SGMARVALITPPIAQMVLSSVYTLAVYSGCCLLAALASCFLEPIETKGGGLQESSHRE 120  
 QY 525 WQGMVGRMGHAGVTRNSGSGQ 548  
 DB 121 WQGMVGRMGHAGVTRNSGSGQ 144

RESULT 6  
 QRYN9  
 ID QRYN9 PRELIMINARY; PRT; 454 AA.  
 AC QRYN9;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 DE SUGAR TRANSPORTER, PUTATIVE.  
 GN DRA0271.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.  
 OX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 radiodurans R1".  
 RL Science 286:1571-1577(1999).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
 DR EMBL; AE001863; AAF12486.1; -;  
 DR TIGR; DRA0271; -;  
 DR InterPro; IPR003662; sub transporter.  
 DR Pfam; PF00083; sugar.tr. 1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
 KW Complete proteome; Sugar transport; Transmembrane.  
 SQ SEQUENCE 454 AA; 48171 MW; 214EA1A3EDC60B88 CRC64;

Query Match 22.6%; Score 644.5; DB 2; Length 454;  
 Best Local Similarity 30.6%; Pred. No. 4.5e-36;  
 Matches 142; Conservative 95; Mismatches 190; Indels 37; Gaps 5;

QY 70 VEDAVEAIGFGKFWKLSVLTGLAWNADAMEMMILSLILAQHCEWRPLPSNOVA--LLTS 127  
 DB 15 VDRVUDDLGLGRFQWKLAIICGLTWAADAMEVLLMGFALPGISAAPFELPKGSPAATMLLT 74  
 QY 128 VVFGMMSSSTLWGNISDOYGRKTKGKISVLWTLVYGLISAFAPVYSWILVRLGLVFGFI 187  
 DB 75 ATFAGMLFGANFWGYLADRVGRRSVFLTTVALGVVFGLAGALAPTLTLLVARFLTGFAI 134  
 QY 188 GG-VPOSVTLVLAELPMKARAKCILLIEVFAIGVVFVVLAVFV---MPSLGNRWLLI 242  
 DB 135 GGTLPVDYSMMAEVPTAWGRFLVYLFESFWAGVIVVVAALAWWYSTAFAPAEGRWLLG 194  
 QY 243 LSAVPLLLFAVLCEFWLPESARYDVLSGNOEKAIATLKRIATENGAMPKLGKLIISRQEDR 302  
 DB 195 LAALPGLVGLIARIIGIPDSPRSLARAGEEAQARAALQKVAQANGGTLTAPLAHPEQPPR 254  
 QY 303 GKMDLFTPHRPWTLLLWFIFWSNAFSYGLV-----LTTTFLFQAGDVCGISSRKNV 357  
 DB 255 VSPAQLFRGVLAARRTPLLMVTWFGLSLGYGIFSWLPSFLRAQGLDLGAV----- 304  
 QY 358 EAKCSLACEYLSEEDYMDLLTWTLLSEFPGLVLTWIIDRLGRKKTMAICFVIFSCSLLL 417  
 DB 305 -----YRSTLLALLAQVPCYLLAAVLEKIGRRVTLVGLTLCGAVGYLF 349  
 QY 418 FICVGRNVLTLLLTARAFISGGFOAAAYVYTPVPTATRALGLGTCSCMARVAGALITPF 477  
 DB 350 LLAHDANTVLTLSALLFALLGANGSLXAYTPELPTPLRTTGMGLVSGVARLASVWSPS 409  
 QY 478 IAOVMLESSVYTLAVYSGCCLLAALASCFLEPIETKGGGLQESS 521  
 DB 410 IGAMLLTGNLTALIVFAVCFALAAALAAWIGVETRGQALAE 453

RESULT 7  
 Q9X1T9  
 ID Q9X1T9 PRELIMINARY; PRT; 422 AA.  
 AC Q9X1T9;  
 DT 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
 DE PERMEASE, PUTATIVE.  
 GN TM1603.  
 OS Thermotoga maritima.  
 OC Bacteria; Thermotogales; Thermotoga.  
 OX NCBI\_TaxID=2336;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MSB8 / DSM 3109;  
 RX MEDLINE=99287316; PubMed=10360571;  
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,  
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
 genome sequence of Thermotoga maritima".  
 RL Nature 399:323-329(1999).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
 DR EMBL; AE001804; AAD36670.1; -;  
 DR TIGR; TM1603; -;  
 DR InterPro; IPR000515; BPD\_transp.  
 DR Pfam; PF00083; sugar.tr. 1.  
 DR PROSITE; PS00402; BPD\_TRANSP\_INN\_MEMBER; UNKNOWN\_1.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; UNKNOWN\_1.  
 KW Complete proteome; Transmembrane.  
 SQ SEQUENCE 422 AA; 47152 MW; 13F9DC5649A1338D CRC64;

Query Match 17.9%; Score 509.5; DB 2; Length 422;  
 Best Local Similarity 29.5%; Pred. No. 5.9e-27;

Matches 135; Conservative 81; Mismatches 188; Indels 53; Gaps 11;

QY 70 VEDAVEAIGFGKFWKLSVLTGLAWMADAMEMMILSILAPOLHCEWRKLPSSQWQVALLTSVV 129  
Db 3 IDIVEIKYVDRKTORRFLITLSIAWDFDAAGVWLLSFVLPYVIKENWLTSTOGATIASAT 62

QY 130 FVGMSSSTLWGNISDOYGRKTKLISVLTWLYYGILSAFAPYVSWILVRLGLVCGIGG 189  
Db 63 FLGMLFCAISVGVADLLGRKVSNNLFFVITFTFLSGFSSFFETLLVRLGLSGFGYGG 122

QY 190 VPOSVTLY-AEFLPMKARAKCILLIEVFWAIGTVFVAVLAVFVMPSLGWRWLLILLSAVPL 248  
Db 123 LMPSEFNAYIAEFTSIRLGRYLVLLSSWAGVSLILGLFVAVNLPN--WRWFVIFSIG- 179

QY 249 LFLAVLCFWLPESARYDVLSSGNOEKATIL-KRIATENGAP-----MPLGKLIISROEDRG 303  
Db 180 YLFVVPFLRMPETPKYAFLLKGGKEALERSLGRVVEEVLPPKKEKVPILALL-----KRE 234

QY 304 KMRDLFTPHFRWTTLLWFTWFSNATSYGLVLLTTLTFOAGDVCGISSRKKAVEAKCSL 363  
Db 235 HLKD-----TVVIWIAWVYVSEVYVIAFTWAPRIESS-----LGVSVYKSS----- 275

QY 364 ACEVLSSEYDMLWTT-----LSEFPGLVLTWIIDRLGRKTKMALCFVIFSPCSLLLFI 419  
Db 276 -----WTFYMWVAQLPGYLSAAFYIEKWKGRKASLGYYFICTGLAALLWAN 321

QY 420 CVGRNVLTLTLTARAFISGGFOAAVYVTPVETATRALGTCGSMARVALITPPIA 479  
Db 322 VRGDASLAAALVSLPFCGLVWGLVAYTPELYPTSLRGTCGGAAGVWARIAGIAPYIT 381

QY 480 QVMLE---SSVYLVAVYSGCLLAALASCFLEPIETKG 514  
Db 382 GFMMKXKSAETLAWISAMFAGVILVIFGRETGK 418

RESULT 8  
Q9HYD9 PRELIMINARY; PRT; 455 AA.  
AC Q9HYD9;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE PROBABLE MFS TRANSPORTER.  
GN PA3467  
OS Pseudomonas aeruginosa  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham Q.-T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goitry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.A., Coulter S.H., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.N., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RA opportunistic pathogen.";  
RT Nature 406:959-964(2000).  
RL -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR ENBL; AE004767; AAG06855.1; -  
DR InterPro: IPR003662; sub.transporter.  
DR Pfam: PF00083; sugar.tr.1  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Complete proteome; Transmembrane.  
SQ SEQUENCE 455 AA; 49357 MW; 2727ED655908A400 CRC64;

Query Match 16.4%; Score 466.5; DB 2; Length 455;

Best Local Similarity 27.7%; Pred. No. 5,3e-24;  
Matches 132; Conservative 94; Mismatches 189; Indels 61; Gaps 13;

QY 72 DAVEAIGFGKFWKLSVLTGLAWMADAMEMMILSILAPOLHCEWRKLPSSQWQVALLTSVV 131  
Db 10 ERLERLPLSPYHRLVPIIALAFFDSMDLAWMTFLGSIKAEFGDLSAQAGLLASSSFF 69

QY 132 GMMSSSTLWGNISDOYGRKTKLIS-VLWTLYYGILSAFAPYVSWILVRLGLVCGIGG-G 189  
Db 70 GMYIGAAISGMLADRFGRKPVFQASIVLWGLASYLCSTAGDLDLS-LTFYVILIGIGMGE 128

QY 190 VPOSVTLY-AEFLPMKARAKCILLIEVFWAIGTVFVAVLAVFVMPSLGWRWLLILLSAVPL 249  
Db 129 FPIAQSLSEMIPASRGKIALMDGFWLPGFVAACLSYFLFLPLTGWRSIFVLALPAV 188

QY 250 LFAVLFCFWLPESARYDVLSSGNOEKATILKRIATE-----NCAPMPILKLIISRQEDR 302  
Db 189 FVLAIRFLIPESPRWLEQAGRRRQADRLVLRDIARVMRSISGLTELPPPL-----ROPQR 242

QY 303 GKNR-----DLFTPHFRWTTLLWFTWFSNATSYGLVLLTTLTFOAGDVCGISSRK 354  
Db 243 ERSRPGFFSFAFELWSPAYRRRTLTVMGLWFFALLGFGYGLTSWLSALLQSGFA----- 296

QY 355 KAVEAKCSLACEVLSSEYDMLWTTLSSEFPGLVLTWIIDRLGRKTKMALCF-----VI 409  
Db 297 -----VTQSYYVTVL- ISLAGIPGFLCAAWLVESWGRKPSCVLMLLGGGAMA 342

QY 410 FSCSLLLFICVGRNVLTLTLF--IARAFISGGFOAAVYVTPVETATRALGTCGSM 467  
Db 343 YAYGQTAVF---GGSLLALIGFLAMQFFLFGMAVLYTYTPELYPTSLRGTCGGAAGV 399

QY 468 ARVGALITPPIAOMLV-----ESSVLTAVYSGCLLAALASCFLEPIETKGGLQE 519  
Db 400 GRIGSLGLPLVTGLVPLTCQGGVFTLGLAL---CFGVAALVWVNAFGIETRGRILEE 452

RESULT 9  
Q9W3W9 PRELIMINARY; PRT; 709 AA.  
AC Q9W3W9;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE CG3168 PROTEIN.  
GN CG3168  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY.  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Chertys J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Munkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,  
 RA Nelson D.R., Nelson K., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of Drosophila melanogaster."  
 RT Science 287:2185-2195(2000).  
 RL Science 287:2185-2195(2000).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
 DR EMBL: AE003438; AAF46193.1; -.  
 DR FlyBase: FBgn0029896; CG3168.  
 DR InterPro: IPR003662; sub.transporter.  
 DR Pfam: PF00083; sugar tr; 1.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
 DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
 KW Transmembrane.  
 SQ SEQUENCE 709 AA; 77930 MW; 1B5AD1E9D133AE94 CRC64;

Query Match 15.7%; Score 446; DB 5; Length 709;  
 Best Local Similarity 25.5%; Pred. No. 2.2e-22;  
 Matches 127; Conservative 98; Mismatches 210; Indels 64; Gaps 12;  
 QY 71 EDVAEAFGRKQKWLVTGLTGLAWMADAMEMMILSILAPQLHCEWRLPSQVALLTSVVF 130  
 DB 215 ERAIELCGYKPHYILLAIAGLSTVSEEMDVISMFLPSAECDLDTETKGLNSLIIF 274  
 QY 131 VGMSSSTLGNISDOYGRKTKLISVLTLYYLGILSAFAPYNSILVRLGLVFGIGV 190  
 DB 275 IGMVYAGVFWGSIADSGRKKVLIIVISFMNAFCIVASSFSQTSYFFMLFRFLNGAALGS 334  
 QY 191 -PQSVTLAEFLPMKARAKCILLIEVFWAIGTVFVWLVAVFVMP-SLG-----WR 238  
 DB 335 GPVWSYFAEFQPKAKRGSMLSFMAAFWTFGNLFVASLAWLIIPRTGTFPTFYNSWR 394  
 QY 239 WLLILSAVPLLLFVLCVLPESARYDVLSGNQEKATLKRIATENGAPPLGLIISR 298  
 DB 395 IFLLVCSPLSFLVGLFLLFPLSPKFLTRGKKDRALAFRGIFVTNKKRPDEYMYVDL 454  
 QY 299 Q-----EDRGKMRDLF-----TPHFRWTTLLLWFIFSNFASYGLVL 336  
 DB 455 EYDEKLLSNGVNNKYSRMISGMVDHSLRPFKSPILRTIVSI-TINFTHIGYGLLM 513  
 QY 337 LTTELF-----QAGDVCGIS-----SRKKAKEAKSLACEYLSSEEDYMDLLW 378  
 DB 514 WPELFNRFEEYEKAFPPQSGACVCAVTVVNLAKESQNNGTCS---SDIQSVEMESLI 570  
 QY 379 TTLSEFPGLVTLWLIIDRLGRKKTKWALCFVIFSCSLILFICVGRNVLTLTLFIARAFIS 438  
 DB 571 SLASALPAMLLAILGMDLGRKFFLIAGTWTAGICSLMYF--VRSSVONL-----VVSATFS 626  
 QY 439 GGFQAAVY-----YTPVPTATRALGLGTCSGMARVGALITPFIQVMSLESSVYLTAVY 494  
 DB 627 GAISAANAALDCLITEVPTKLRATGVAISMVAARLGGIIGNIVIAQLLDNVCPSPTFIV 686  
 QY 495 SCCCLLAALASCLFIEFK 513  
 DB 687 SGLLGGGLMCLLLPNTTR 705  
 RESULT 10

P70485  
 ID P70485 PRELIMINARY; PRT; 593 AA.  
 AC Q9ROW2  
 DT 01-FEB-1997 (Tremblrel. 02, Created)  
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE ORGANIC CATION TRANSPORTER OCT2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;  
 RX MEDLINE=96295517; PubMed=8702418;  
 RA Okuda M., Saito H., Urakami Y., Takano M., Inui K.;  
 RT "cDNA cloning and functional expression of a novel rat kidney organic  
 RT cation transporter, OCT2."  
 RL Biochem. Biophys. Res. Commun. 224:500-507(1996).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
 DR EMBL: D83044; BA011754.1; -.  
 DR InterPro: IPR003662; sub.transporter.  
 DR Pfam: PF00083; sugar tr; 1.  
 DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_2.  
 KW Transmembrane.  
 SQ SEQUENCE 593 AA; 66080 MW; 36C1044E0C04B443 CRC64;  
 Query Match 15.3%; Score 436; DB 11; Length 593;  
 Best Local Similarity 29.5%; Pred. No. 8.4e-22;  
 Matches 127; Conservative 77; Mismatches 148; Indels 78; Gaps 16;  
 QY 119 SNOVALITVVPVGMSSSTLGNISDOYGRKTKLISVLTLYYLGILSAFAPYNSILV 178  
 DB 146 SMLDLQFQSVWVNGFFIGAMMIGYLADRFGRKFLVTLINAIAGMALMAISPNYAMLV 205  
 QY 179 ---LRGLV---GFGIGGVPSVTLVYAEFLPMKARAKCILLIEVFWAIGTVFVWLVLA--VF 230  
 DB 206 FRFLOGLVSKAGWLGVI-----LITEFVGLGRVWVGICVQIAFTVGL--LILAGVAY 257  
 QY 231 VMPSLGWRMLLILSAVPLLLFAVLGFW-LPESARYDVLSGNQEKATLKRIATENGAPM 289  
 DB 258 VIPN--WRWLQFAVTLPNFCF--LLYFWCIPESPRWLLISQNKIVKAMKIIKHIAKNGKSV 314  
 QY 290 PLGKLIIISQEDRGK-----MRDLF--TPHFRWTTLLLWFIFSNFASYGLVLTTELFQ 343  
 DB 315 PVSQNLTLDPEDAGKRLKPSIDLVRTPOIRKHTLLIMNWTSSVLYOGLIM---HMGL 371  
 QY 344 AGDVCGIISRRKKAVEAKCSLACEYLSSEEDYMDLLWTLTSEFPGLVTLWIIDRLGRKKT 403  
 DB 372 AGD-----NIYLDFFYSALVEFFPAAFIILITIDRVGRRPW 407  
 QY 404 ALC-FWIFSCSLILFI-----CVGRNVLTLLLFIAARAFISGFGQAAVYVTP 450  
 DB 408 AVSNVAGAACLASVFIIPDDLQMLKTIACLGRMGITM-----AYEMVCLVNAE 456  
 QY 451 VYPTATRALGLGTCSGMARVGALITPFIQVMSLESSVYLTAVYSGCCLLAALASCLFLPI 510  
 DB 457 LYPTIYRNLGVLCVSSMCDIGIITPFLVYRLTDIMPEPLVVFVAVGLVAGALLLP- 515  
 QY 511 ETKGGGLQES 520  
 DB 516 ETKGRALPET 525  
 RESULT 11  
 ID Q9ROW2 PRELIMINARY; PRT; 555 AA.  
 AC Q9ROW2  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE	ORGANIC CATION TRANSPORTER OCT2R.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=KIDNEY;
RA	MEDLINE=99316020; PubMed=10385678;
RR	Grudemann D., Liebich G., Kiefer N., Koester S., Schoemig E.;
RT	"Selective substrates for non-neuronal monoamine transporters.";
RL	Mol. Pharmacol. 56:1-10(1999).
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC	-!- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.
EMBL:	YI3154; CAB52215.1; -.
DR	InterPro: IPR003662; sub.transporter.
DR	Pfam: PF00083; sugar_tr.1.
DR	PROSITE: PS00216; SUGAR_TRANSPORT_1; UNKNOWN_2.
KW	Transmembrane.
SQ	SEQUENCE 555 AA; 62342 MW; 29521969AELAC206 CRC64;
Query Match	15.3%; Score 435; DB 11; Length 555;
Best Local Similarity	29.5%; Pred. No. 9,1e-22;
Matches 127; Conservative 76; Mismatches 149; Indels 78; Gaps	
QY	119 SHQVALLTSVFVGMMSSSTLNGNISDOYCRKTGLKISLVLTLYYLGSFAFVYSMLVL 178             : : :         :
DB	146 SWMLDFQSVDVVNPGFFIGAMMIGYLADRGKFCLLVTILINISGALMAISPNTAAWLVL 205         :     :     :     :     :     :     :     :     :
QY	179 ---LRGLV---GFGIGGVPSQSVTLXAEFLPMKARAKCILLIEVFVAIGTVFEVLA--VF 230         :     :     :     :     :     :     :     :     :
DB	206 FRELQGLVSKAGWLIGYI-----LITEFVLGYRRVMVGICYQIAFTVGL---LILAGVAY 257         :     :     :     :     :     :     :     :     :
QY	231 VMSLGWRWLLLSAVPLLLFVLCFW-LPESARYDVLDSQEKATATLKRIATENGAPM 289 
DB	258 VFEN--WRWLOFATVLPNCF-LLYFMCIPESPRLWISONKI VKAMKIITKHIAKKGKSV 314 
QY	290 PLCKLIISRODRGK-----MRDLF-TPHFRWTLLWFIFWSNAFYSGVLVLLTLELQ 343 
DB	315 PVSJLNLTPDEDAGKLNSFDLDVETPOIRKHTLLIMYNWFTSSVLYOGLM---HMGL 371 
QY	344 AGDVCGISSRKKAWEAKCSLAYEYLSEEDYMOLLWTTLSFEFGVGLVTLMIIDRLGRKKTM 403 
DB	372 AGD-----NIYLDFFYSALVEFFAAFIITLIDRVGRPYM 407 
QY	404 ALC-FVIESPCSLLFPI-----CVGNRLVTLLFLFIARAFISGGFOAAVYVTP 450 
DB	408 AVSNWAGAACLASVIPDDLQWLKITIACLRMGITM-----AYEMVCLVNAE 456 
QY	451 VYPTATRALGLGTSCGMARVGALITPFIIAQVMLESSYYLTVLAVYSGCCLLAALASCFLPI 510 
DB	457 LYPTYIRNLGVLCSSMCDGGIITPFLVYRTDIWMERPLVVFVAVGLVAGALVLLLP- 515 
QY	511 ETKGGGLOES 520
DB	516 ETGKKALPET 525
RESULT 12	
P97558	
ID	P97558 PRELIMINARY; PRT; 593 AA.
AC	D
DC	A97558;
DT	01-MAY-1997 (TrEMBLrel. 03, Created)
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	ORGANIC CATION TRANSPORTER.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;	[1]
RP	SEQUENCE FROM N.A.

[illegible]

Wed Mar 13 13:08:17 2002

DR InterPro; IPR003662; sub\_transporter.  
DR Pfam; PF00083; sugar\_tr; 1.  
KW Transmembrane.  
SQ SEQUENCE 535 AA; 58712 MW; E5C189

Query Match 15.2%; Score 432; DB 11; Length 535;  
Best Local Similarity 28.7%; Pred. No. 1.e-21;  
Matches 149; Conservative 79; Mismatches 215; Indels 7

QY	34	HEVQIEGVHVGLEAVEIDDGAAPKFEANPTDDTETM-----VEDAVEAIGF	80
Db	47	HHCALPGAPANILSHODLWLEAHLPRE---TDGSSSCLREAFYPOIVPNTVLGTETVNSG	102
QY	81	KFQWK-LSVLTGLTGLAWMADAMEMMILSILAPLQHCERHRLPSQVAL--LTSV-VFVCMKSS	136
Db	103	EPEGEPLTVPSCSGWEYDRSE-----FSSITATEMDLVCQORGLNKTITSTCFFIGVLVG	156
QY	137	STLWGNISDOYGRKTKGLISVLTLYXGLISAFAPYISWILVLRLGVLGFGIGGVPOSQ-V	195
Db	157	AVVGYLSDRFGRRRLLVAVVSSLVGLMLSAASINYMVTVTRTUTGSALAGFTIILVP	216
QY	196	LYAEFLPMKARAKCILLIEVFNAIGTVFEVVLAVFVWPSLQWHPWLLLSAVPLLLFVLC	255
Db	217	LELEWDOEHRTVAGVISTVFWSS-GGVLLALVUGYLIIRS--WRWLLLAATLPCVCGIISI	273
QY	256	FWLPE\$ARYDVL\$GNO\$KATIKRIATENGAPMPL\$KL-----IISROEDR\$KMR	306
Db	274	WVVPESARWLLTQGRVEAKKYL\$LL\$CAK\$INGRPVGE\$LSOEALNNVVTMERALQRP\$YL	333
QY	307	DLP-TPIHRTWTL\$WIFW\$SNAF\$Y\$GLV\$LLLT\$ELFQAGDVCGIS\$R\$K\$KAVEAK\$CLAC	365
Db	334	DLFRT\$QLR\$H\$IL\$C\$M\$W\$F\$GVN\$F\$Y\$GL\$FL-----DV\$G-----	368
QY	366	EVLSEEDYMDL\$WTLT\$SEFFPGVLVTLWIDRLGRK\$KTM\$--LC\$FVIF\$F\$C\$LL\$LFICVGR	423
Db	369	--L\$LN\$VYQ\$TOLL\$F\$AVE\$LP\$K\$IN\$VY\$FLVR\$R\$R\$RLTEAG\$ML\$G\$A\$LLT\$F\$G\$T\$S\$LV\$S\$LETK	426
QY	424	NVLTL\$LL\$F\$IARAF\$IS\$G\$FOA\$Y\$V\$V\$T\$PEY\$T\$AT\$F\$AL\$G\$TC\$S\$G\$M\$AR\$V\$G\$AL\$IT\$F\$IAQ\$V\$ML	483
Db	427	SWITALVVG\$G\$F\$E\$A\$A\$F\$T\$T\$A\$Y\$L\$T\$E\$S\$E\$P\$TV\$V\$Q\$T\$G\$IG\$T\$AL\$M\$GR\$G\$AS\$LAR\$A\$A\$---L	483
QY	484	ESSVY\$TL\$--AVY\$G\$C\$LLA\$A\$AL\$CF\$PI\$ET\$K\$G\$G\$LO\$E\$	520
Db	484	LDG\$W\$LL\$PK\$VAY\$G\$G\$IAL\$V\$A\$A\$CT\$ALL\$P\$-ET\$K\$A\$OL\$P\$ET	521

RESULT 14  
Q63089  
ID Q63080  
PRELIMINARY.  
PRT: 556 AA.

AC	Q63089;
AD	FAHLMANNR1;
AE	
AF	
AG	
AH	
AI	
AJ	
AK	
AL	
AM	
AN	
AO	
AP	
AQ	
AR	
AS	
AT	
AV	
AW	
AX	
AY	
AZ	
BA	
BB	
BC	
BD	
BE	
BF	
BG	
BH	
BI	
BJ	
BK	
BL	
BM	
BN	
BO	
BP	
BQ	
BR	
BS	
BT	
BU	
BV	
BW	
BX	
BY	
BZ	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	
CX	
CY	
CA	
CB	
CC	
CD	
CE	
CF	
CG	
CH	
CI	
CJ	
CK	
CL	
CM	
CN	
CO	
CP	
CQ	
CR	
CS	
CT	
CU	
CV	
CW	

RP SEQUENCE FROM N. A.  
RC TISSUE=LIVER;  
RA Gorboulev V.G.;  
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.  
DR EMBL: X78855; CAA55411.1; -  
DR InterPro: IPR003662; sub.transporter.  
DR Pfam: PF00083; sugar\_tr\_1  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_1.  
KW Transmembrane.  
SQ SEQUENCE 556 AA; 61541 MW; 9F42131CCCEC0920 CRC64;

Query Match 15.1%; Score 429.5; DB 11; Length 556;  
Best Local Similarity 29.3%; Pred. No. 2.2e-21;  
Matches 127; Conservative 76; Mismatches 166; Indels 65; Gaps 12;

[illegible]

RESULT	15	
002713		
ID	002713	
DC	002713	PRELIMINARY;
DT	03-JUL-1997	04, Created
DT	01-JUL-1997	(TREMBLUREL 04, Last sequence update)
DT	01-JUN-2001	(TREMBLUREL 17, Last annotation update)
DE	APRCA	OPTICAL CATION TRANSPORTER.

OCT-2P.  
 Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OC NCBI\_TaxId=9823;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC TISSUE=KIDNEY;  
 RX MEDLINE=97256752; PubMed=9099681;  
 RA Gruendemann D., Sabin-Ebell J., Martel F., Oerding N., Schmidt A.  
 RA Schlemmig E.;  
 RT "Primary structure and functional expression of the apical organ-  
 cation transporter from kidney epithelial LLC-PK1 cells.;"  
 RL J. Biol. Chem. 272:10408-10413(1997).

Wed Mar 13 13:08:17 2002

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: TO SUGAR TRANSPORTER FAMILY.

CC EMBL: Y09400; CAA70567.1; -  
DR InterPro: IPR003662; sub-transporter.  
DR Pfam: PF00083; sugar tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN\_2.  
DR Transmembrane.  
KW SEQUENCE 554 AA; 61989 MW; E14B5565600C553B CRC64;

Query Match 15.1%; Score 429; DB 6; Length 554;  
Best Local Similarity 30.3%; Pred. No. 2.3e-21;  
Matches 130; Conservative 65; Mismatches 158; Indels 76; Gaps 14;

119	QY	SNQVALLTSVVFVGMSSSTLWGNISDQYGRKTKGISVLTWLTLYGYILSAFAPVYSWILV	178
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
145	Db	SWLLDLFQSAVNVGFFIGSVGIGYIADRFGRKLCLLLTILINAVSGVLMAISPTVTMWLV	204
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
179	QY	LR---GLV---GFGIGGVQSVTLVYAEFLPMKARAKCILLIEVFVAIGVVFVFLA--VF	230
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
205	Db	FRLIOGLVSKAGNMIGYI-----LITEFVGLSVRRVTGVIFYQVATFGL--LVLAGVAY	256
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
231	QY	VMPSLGWRWLLIILSAVPLLLFAVLCPWLPESARYDVLSONQEKAIATLKRIATENACMP	290
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
257	Db	ALPH--WRWLQFTVTLPNFCFLFYWCVPESPRESPLLISQNKAKAMSIKHIKAKNGKSLP	314
		:  :	
291	QY	LGKLIISRQEDRG--KMRDLF-----TPHPRWTLLILFWTFESNAFSYVGLVLTTELFOA	344
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
315	Db	ASLOSRLAPDEVEGKLPSPFLDLVTRTPQIRKHTLTLIMYNTFESAVLYQGLVM-----	366
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
345	QY	GDVCGISSRKKAVEAKCSLACEYLSEEDYMDLLWTTLSEFPFGVLVTLTWIIDRLGRKKTMA	404
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
367	Db	-----HMGLAGSNL-----YLDFFYSALVEPPAALLILLTIDRLGRHRPWA	407
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
405	QY	LCFVIF--SFGSLLLFYI-----CVGRNVLTLLFLIARAFISGGFOAAVYVTPV	451
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
408	Db	ASNVVAGAACLASVFIPEDPHHLIRITVCLUGRMGITM-----AYENVCUNNAEL	456
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
452	QY	YPTATRALGLGTCSGMARVGALLTPFIAQVMELESSVYLTIAVYSGCCLLAALASCPLPIE	511
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
457	Db	YPTFIRNLGVLVCSMCDIGGIITPFVLYRLTDIHWHELPVAVFVAVGLIAGGLVLLLP-E	515
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
512	QY	TKGGGLQES	520
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	
516	Db	TKGKTLPET	524
		: : :    : : :    : : :    : : :    : : :    : : :    : : :    : : :	

Search completed: March 13, 2002, 12:42:29  
Job time: 135 sec